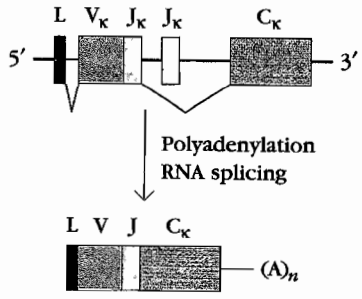
chapter 5

Organization and Expression of Immunoglobulin Genes



Kappa Light-Chain Gene Rearrangement

ONE OF THE MOST REMARKABLE FEATURES OF the vertebrate immune system is its ability to respond to an apparently limitless array of foreign antigens. As immunoglobulin (Ig) sequence data accumulated, virtually every antibody molecule studied was found to contain a unique amino acid sequence in its variable region but only one of a limited number of invariant sequences in its constant region. The genetic basis for this combination of constancy and tremendous variation in a single protein molecule lies in the organization of the immunoglobulin genes.

In germ-line DNA, multiple gene segments encode portions of a single immunoglobulin heavy or light chain. These gene segments are carried in the germ cells but cannot be transcribed and translated into complete chains until they are rearranged into functional genes. During B-cell maturation in the bone marrow, certain of these gene segments are randomly shuffled by a dynamic genetic system capable of generating more than 10^6 combinations. Subsequent processes increase the diversity of the repertoire of antibody binding sites to a very large number that exceeds 10^6 by at least two or three orders of magnitude. The processes of B-cell development are carefully regulated: the maturation of a progenitor B cell progresses through an ordered sequence of Ig-gene rearrangements, coupled with modifications to the gene that contribute to the diversity of the final product. By the end of this process, a mature, immunocompetent B cell will contain coding sequences for one functional heavy-chain variable-region and one light-chain variable-region. The individual B cell is thus antigenically committed to a specific epitope. After antigenic stimulation of a mature B cell in peripheral lymphoid organs, further rearrangement of constant-region gene segments can generate changes in the isotype expressed, which produce changes in the biological effector functions of the immunoglobulin molecule without changing its specificity. Thus, mature B cells contain chromosomal DNA that is no longer identical to germ-line DNA. While we think of genomic DNA as a stable genetic blueprint, the lymphocyte cell lineage does not retain an intact copy of this blueprint. Genomic rearrangement is an essential feature of lymphocyte differentiation, and no other vertebrate cell type has been shown to undergo this process.

This chapter first describes the detailed organization of the immunoglobulin genes, the process of Ig-gene rearrangement, and the mechanisms by which the dynamic immunoglobulin genetic system generates more than 10^8 different antigenic specificities. Then it describes the mechanism of class switching, the role of differential RNA processing in the expression of immunoglobulin genes, and the regulation of Ig-gene transcription. The chapter concludes with the application of our knowledge of the molecular

VISUALIZING CONCEPTS

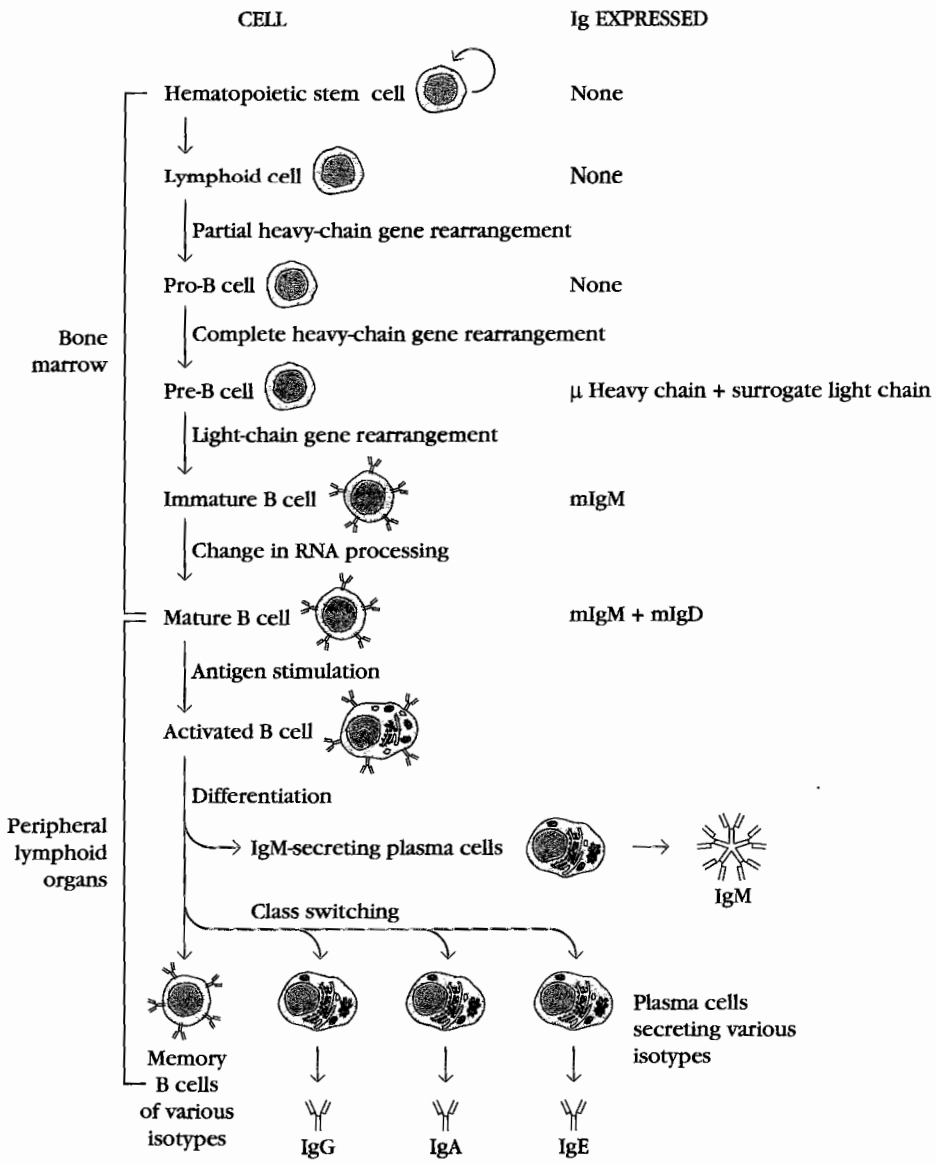


FIGURE 5-1 Overview of B-cell development. The events that occur during maturation in the bone marrow do not require antigen, whereas activation and differentiation of mature B cells in peripheral lymphoid organs require antigen. The labels mIgM and mIgD refer to membrane-associated Igs. IgG, IgA, and IgE are secreted immunoglobulins.

biology of immunoglobulin genes to the engineering of antibody molecules for therapeutic and research applications. Chapter 11 covers in detail the entire process of B-cell development from the first gene rearrangements in progenitor B cells to final differentiation into memory B cells and antibody-secreting plasma cells. Figure 5-1 outlines the sequential stages in B-cell development, many of which result from critical rearrangements.

Genetic Model Compatible with Ig Structure

The results of the immunoglobulin-sequencing studies described in Chapter 4 revealed a number of features of immunoglobulin structure that were difficult to reconcile with classic genetic models. Any viable model of the

immunoglobulin genes had to account for the following properties of antibodies:

- The vast diversity of antibody specificities
- The presence in Ig heavy and light chains of a variable region at the amino-terminal end and a constant region at the carboxyl-terminal end
- The existence of isotypes with the same antigenic specificity, which result from the association of a given variable region with different heavy-chain constant regions

Germ-Line and Somatic-Variation Models Contended to Explain Antibody Diversity

For several decades, immunologists sought to imagine a genetic mechanism that could explain the tremendous diversity of antibody structure. Two different sets of theories emerged. The **germ-line theories** maintained that the genome contributed by the germ cells, egg and sperm, contains a large repertoire of immunoglobulin genes; thus, these theories invoked no special genetic mechanisms to account for antibody diversity. They argued that the immense survival value of the immune system justified the dedication of a significant fraction of the genome to the coding of antibodies. In contrast, the **somatic-variation theories** maintained that the genome contains a relatively small number of immunoglobulin genes, from which a large number of antibody specificities are generated in the somatic cells by mutation or recombination.

As the amino acid sequences of more and more immunoglobulins were determined, it became clear that there must be mechanisms not only for generating antibody diversity but also for maintaining constancy. Whether diversity was generated by germ-line or by somatic mechanisms, a paradox remained: How could stability be maintained in the constant (C) region while some kind of diversifying mechanism generated the variable (V) region?

Neither the germ-line nor the somatic-variation proponents could offer a reasonable explanation for this central feature of immunoglobulin structure. Germ-line proponents found it difficult to account for an evolutionary mechanism that could generate diversity in the variable part of the many heavy- and light-chain genes while preserving the constant region of each unchanged. Somatic-variation proponents found it difficult to conceive of a mechanism that could diversify the variable region of a single heavy- or light-chain gene in the somatic cells without allowing alteration in the amino acid sequence encoded by the constant region.

A third structural feature requiring an explanation emerged when amino acid sequencing of the human myeloma protein called Til revealed that identical variable-region sequences were associated with both γ and μ heavy-chain constant regions. A similar phenomenon was observed in rabbits by C. Todd, who found that a particular allotypic marker in the heavy-chain variable region could be associated with α, γ, and μ heavy-chain constant regions. Considerable additional evidence has confirmed that a single variable-region sequence, defining a particular antigenic specificity, can be associated with multiple heavy-chain constant-region sequences; in other words, different classes, or isotypes, of antibody (e.g., IgG, IgM) can be expressed with identical variable-region sequences.

Dreyer and Bennett Proposed the Two-Gene Model

In an attempt to develop a genetic model consistent with the known findings about the structure of immunoglobulins, W. Dreyer and J. Bennett suggested, in their classic theoretical paper of 1965, that two separate genes encode a single immunoglobulin heavy or light chain, one gene for the V region (variable region) and the other for the C region (constant region). They suggested that these two genes must somehow come together at the DNA level to form a continuous message that can be transcribed and translated into a single Ig heavy or light chain. Moreover, they proposed that hundreds or thousands of V-region genes were carried in the germ line, whereas only single copies of C-region class and subclass genes need exist.

The strength of this type of recombinational model (which combined elements of the germ-line and somatic-variation theories) was that it could account for those immunoglobulins in which a single V region was combined with various C regions. By postulating a single constant-region gene for each immunoglobulin class and subclass, the model also could account for the conservation of necessary biological effector functions while allowing for evolutionary diversification of variable-region genes.

At first, support for the Dreyer and Bennett hypothesis was indirect. Early studies of DNA hybridization kinetics using a radioactive constant-region DNA probe indicated that the probe hybridized with only one or two genes, confirming the model's prediction that only one or two copies of each constant-region class and subclass gene existed. However, indirect evidence was not enough to overcome stubborn resistance in the scientific community to the hypothesis of Dreyer and Bennet. The suggestion that two genes encoded a single polypeptide contradicted the existing one gene–one polypeptide principle and was without precedent in any known biological system.

As so often is the case in science, theoretical and intellectual understanding of Ig-gene organization progressed ahead of the available methodology. Although the Dreyer and Bennett model provided a theoretical framework for reconciling the dilemma between Ig-sequence data and gene organization, actual validation of their hypothesis had to wait for several major technological advances in the field of molecular biology.

Tonegawa's Bombshell—Immunoglobulin Genes Rearrange

In 1976, S. Tonegawa and N. Hozumi found the first direct evidence that separate genes encode the V and C regions of immunoglobulins and that the genes are rearranged in the course of B-cell differentiation. This work changed the field of immunology. In 1987, Tonegawa was awarded the Nobel Prize for this work.

Selecting DNA from embryonic cells and adult myeloma cells—cells at widely different stages of development—Tonegawa and Hozumi used various restriction endonucleases to generate DNA fragments. The fragments were then separated by size and analyzed for their ability to hybridize with a radiolabeled mRNA probe. Two separate restriction fragments from the embryonic DNA hybridized with the mRNA, whereas only a single restriction fragment of the adult myeloma DNA hybridized with the same probe. Tonegawa and Hozumi suggested that, during differentiation of lymphocytes from the embryonic state to the fully differentiated plasma-cell stage (represented in their system by the myeloma cells), the V and C genes undergo rearrangement. In the embryo, the V and C genes are separated by a large DNA segment that contains a restriction-endonuclease site; during differentiation, the V and C genes are brought closer together and the intervening DNA sequence is eliminated.

The pioneering experiments of Tonegawa and Hozumi employed a tedious and time-consuming procedure that has since been replaced by the much more powerful approach of Southern-blot analysis. This method, now universally used to investigate the rearrangement of immunoglobulin genes, eliminates the need to elute the separated DNA restriction fragments from gel slices prior to analysis by hybridization with an immunoglobulin gene segment probe. Figure 5-2 shows the detection of rearrangement at the κ light-chain locus by comparing the fragments produced by digestion of DNA from a clone of B-lineage cells with the pattern obtained by digestion of non-B cells (e.g., sperm or liver cells). The rearrangement of a V gene deletes an extensive section of germ-line DNA, thereby creating differences between rearranged and unrearranged Ig loci in the distribution and number of restriction sites. This results in the generation of

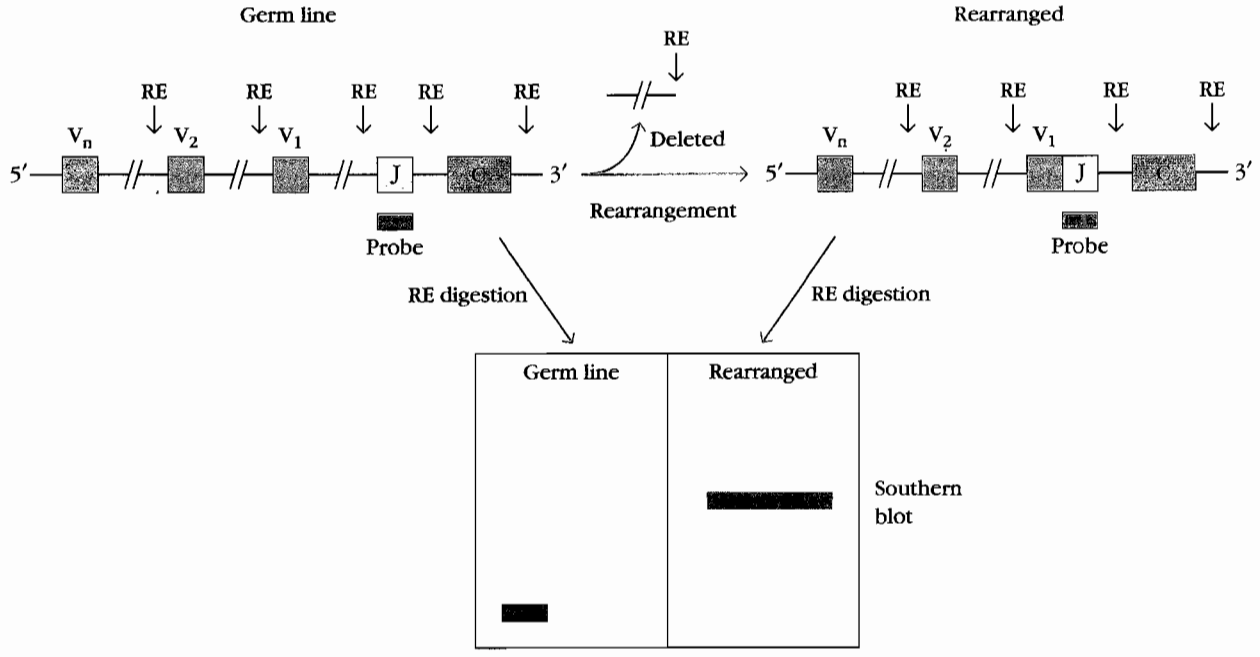


FIGURE 5-2 Experimental basis for diagnosis of rearrangement at an immunoglobulin locus. The number and size of restriction fragments generated by the treatment of DNA with a restriction enzyme is determined by the sequence of the DNA.The digestion of rearranged DNA with a restriction enzyme (RE) yields a pattern of restriction fragments that differ from those obtained by digestion of an unrearranged locus with the same RE. Typically, the fragments are analyzed by the technique of Southern blotting. In this example, a probe that includes a J gene segment is used to identify RE digestion fragments that include all or portions of this segment. As shown, rearrangement results in the deletion of a segment of germ-line DNA and the loss of the restriction sites that it includes. It also results in the joining of gene segments, in this case a V and a J segment, that are separated in the germ line. Consequently, fragments dependent on the presence of this segment for their generation are absent from the restriction-enzyme digest of DNA from the rearranged locus. Furthermore, rearranged DNA gives rise to novel fragments that are absent from digests of DNA in the germ-line configuration. This can be useful because both B cells and non-B cells have two immunoglobulin loci. One of these is rearranged and the other is not. Consequently, unless a genetic accident has resulted in the loss of the germ-line locus, digestion of DNA from a myeloma or normal B-cell clone will produce a pattern of restriction that includes all of those in a germ-line digest plus any novel fragments that are generated from the change in DNA sequence that accompanies rearrangement. Note that only one of the several J gene segements present is shown.

different restriction patterns by rearranged and unrearranged loci. Extensive application of this approach has demonstrated that the Dreyer and Bennett two-gene model—one gene encoding the variable region and another encoding the constant region—applied to both heavy and light-chain genes.

Multigene Organization of Ig Genes

As cloning and sequencing of the light- and heavy-chain DNA was accomplished, even greater complexity was revealed than had been predicted by Dreyer and Bennett. The κ and λ light chains and the heavy chains are encoded by separate multigene families situated on different chromosomes (Table 5-1). In germ-line DNA, each of these multigene families contains several coding sequences, called **gene segments,** separated by noncoding regions. During B-cell maturation, these gene segments are rearranged and brought together to form functional immunoglobulin genes.

Each Multigene Family Has Distinct Features

The κ and λ light-chain families contain **V, J,** and **C gene segments;** the rearranged VJ segments encode the variable region of the light chains. The heavy-chain family contains **V, D, J,** and **C gene segments;** the rearranged VDJ gene segments encode the variable region of the heavy chain. In each gene family, C gene segments encode the constant regions. Each V gene segment is preceded at its 5′ end by a small exon that encodes a short **signal** or **leader (L) peptide** that guides the heavy or light chain through the endoplasmic reticulum. The signal peptide is cleaved from the nascent light and heavy chains before assembly of the finished immunoglobulin molecule. Thus, amino acids encoded by this leader sequence do not appear in the immunoglobulin molecule.

λ-CHAIN MULTIGENE FAMILY

The first evidence that the light-chain variable region was actually encoded by two gene segments appeared when Tonegawa cloned the germ-line DNA that encodes the variable region of mouse λ light chain and determined its complete nucleotide sequence. When the nucleotide sequence was compared with the known amino acid sequence of the λ-chain variable region, an unusual discrepancy was observed. Although the first 97 amino acids of the λ-chain variable region corresponded to the nucleotide codon sequence, the remaining 13 carboxyl-terminal amino acids of the protein's variable region did not. It turned out that many base pairs away a separate, 39-bp gene segment, called J for *joining,* encoded the remaining 13 amino acids of the λ-chain variable region. Thus, a functional λ variable-region gene contains two coding segments—a 5′ V segment and a 3′ J segment—which are separated by a noncoding DNA sequence in unrearranged germ-line DNA.

The λ multigene family in the mouse germ line contains three V_λ gene segments, four J_λ gene segments, and four C_λ gene segments (Figure 5-3a). The $J_\lambda 4$ is a **pseudogene,** a defective gene that is incapable of encoding protein; such genes are indicated with the psi symbol (ψ). Interestingly, $J_\lambda 4$'s constant region partner, $C_\lambda 4$, is a perfectly functional gene. The V_λ and the three functional J_λ gene segments encode the variable region of the light chain, and each of the three functional C_λ gene segments encodes the constant region of one of the three λ-chain subtypes (λ1, λ2, and λ3). In humans, the lambda locus is more complex. There are 31 functional V_λ gene segments, 4 J_λ segments, and 7 C_λ segments. In additional to the functional gene segments, the human lambda complex contains many V_λ, J_λ, and C_λ pseudogenes.

κ-CHAIN MULTIGENE FAMILY

The κ-chain multigene family in the mouse contains approximately 85 V_κ gene segments, each with an adjacent leader sequence a short distance upstream (i.e., on the 5′ side). There are five J_κ gene segments (one of which is a nonfunctional pseudogene) and a single C_κ gene segment (Figure 5-3b). As in the λ multigene family, the V_κ and J_κ gene segments encode the variable region of the κ light chain, and the C_κ gene segment encodes the constant region. Since there is only one C_κ gene segment, there are no subtypes of κ light chains. Comparison of parts *a* and *b* of Figure 5-3 shows that the arrangement of the gene segments is quite different in the κ and λ gene families. The κ-chain multigene family in humans, which has an organization similar to that of the mouse, contains approximately 40 V_κ gene segments, 5 J_κ segments, and a single C_κ segment.

HEAVY-CHAIN MULTIGENE FAMILY

The organization of the immunoglobulin heavy-chain genes is similar to, but more complex than, that of the κ and λ light-chain genes (Figure 5-3c). An additional gene segment encodes part of the heavy-chain variable region. The existence of this gene segment was first proposed by Leroy Hood and his colleagues, who compared the heavy-chain variable-region amino acid sequence with the V_H and J_H nucleotide sequences. The V_H gene segment was found to encode amino acids 1 to 94 and the J_H gene segment

TABLE 5-1 Chromosomal locations of immunoglobulin genes in human and mouse

Gene	CHROMOSOME	
	Human	Mouse
λ Light chain	22	16
κ Light chain	2	6
Heavy chain	14	12

VISUALIZING CONCEPTS

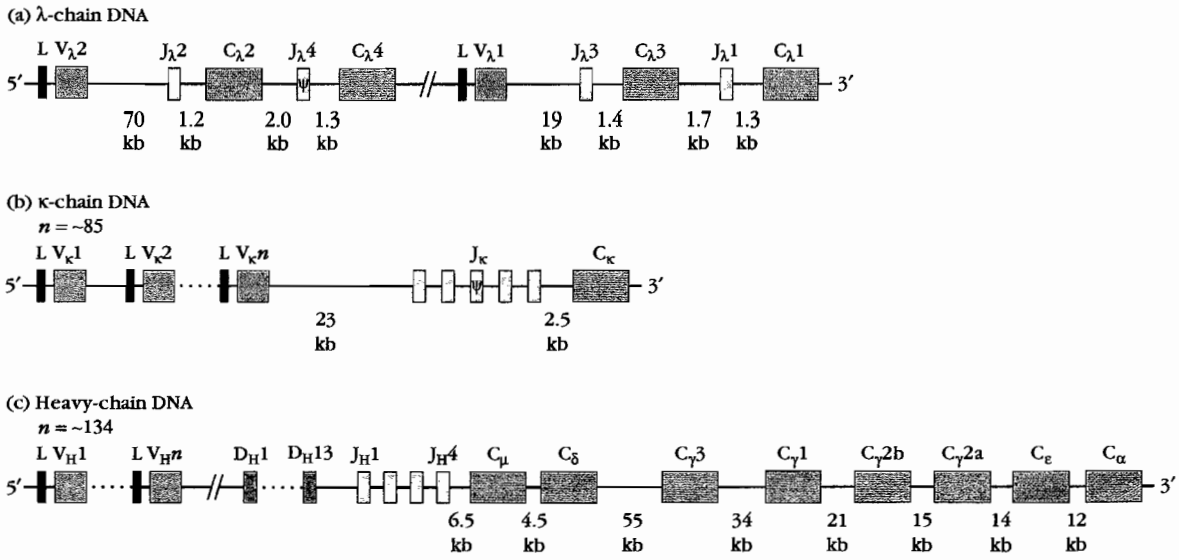


FIGURE 5-3 Organization of immunoglobulin germ-line gene segments in the mouse: (a) λ light chain, (b) κ light chain, and (c) heavy chain. The λ and κ light chains are encoded by V, J, and C gene segments. The heavy chain is encoded by V, D, J, and C gene segments. The distances in kilobases (kb) separating the various gene segments in mouse germ-line DNA are shown below each chain diagram.

was found to encode amino acids 98 to 113; however, neither of these gene segments carried the information to encode amino acids 95 to 97. When the nucleotide sequence was determined for a rearranged myeloma DNA and compared with the germ-line DNA sequence, an additional nucleotide sequence was observed between the V_H and J_H gene segments. This nucleotide sequence corresponded to amino acids 95 to 97 of the heavy chain.

From these results, Hood and his colleagues proposed that a third germ-line gene segment must join with the V_H and J_H gene segments to encode the entire variable region of the heavy chain. This gene segment, which encoded amino acids within the third complementarity-determining region (CDR3), was designated D for *diversity*, because of its contribution to the generation of antibody diversity. Tonegawa and his colleagues located the D gene segments within mouse germ-line DNA with a cDNA probe complementary to the D region, which hybridized with a stretch of DNA lying between the V_H and J_H gene segments.

The heavy-chain multigene family on human chromosome 14 has been shown by direct sequencing of DNA to contain 51 V_H gene segments located upstream from a cluster of 27 functional D_H gene segments. As with the light-chain genes, each V_H gene segment is preceded by a leader sequence a short distance upstream. Downstream from the D_H gene segments are six functional J_H gene segments, followed by a series of C_H gene segments. Each C_H gene segment encodes the constant region of an immunoglobulin heavy-chain isotype. The C_H gene segments consist of coding exons and noncoding introns. Each exon encodes a separate domain of the heavy-chain constant region. A similar heavy-chain gene organization is found in the mouse.

The conservation of important biological effector functions of the antibody molecule is maintained by the limited number of heavy-chain constant-region genes. In humans and mice, the C_H gene segments are arranged sequentially in the order C_μ, C_δ, C_γ, C_ε, C_α (see Figure 5-3c). This sequential arrangement is no accident; it is generally related to the sequential expression of the immunoglobulin classes in the course of B-cell development and the initial IgM response of a B cell to its first encounter with an antigen.

Variable-Region Gene Rearrangements

The preceding sections have shown that functional genes that encode immunoglobulin light and heavy chains are

assembled by recombinational events at the DNA level. These events and the parallel events involving T-receptor genes are the only known site-specific DNA rearrangements in vertebrates. Variable-region gene rearrangements occur in an ordered sequence during B-cell maturation in the bone marrow. The heavy-chain variable-region genes rearrange first, then the light-chain variable-region genes. At the end of this process, each B cell contains a single functional variable-region DNA sequence for its heavy chain and another for its light chain.

The process of variable-region gene rearrangement produces mature, immunocompetent B cells; each such cell is committed to produce antibody with a binding site encoded by the particular sequence of its rearranged V genes. As described later in this chapter, rearrangements of the heavy-chain constant-region genes will generate further changes in the immunoglobulin class (isotype) expressed by a B cell, but those changes will not affect the cell's antigenic specificity.

The steps in variable-region gene rearrangement occur in an ordered sequence, but they are random events that result in the random determination of B-cell specificity. The order, mechanism, and consequences of these rearrangements are described in this section.

Light-Chain DNA Undergoes V-J Rearrangements

Expression of both κ and λ light chains requires rearrangement of the variable-region V and J gene segments. In humans, any of the functional V_λ genes can combine with any of the four functional J_λ-C_λ combinations. In the mouse, things are slightly more complicated. DNA rearrangement can join the $V_\lambda 1$ gene segment with either the $J_\lambda 1$ or the $J_\lambda 3$ gene segment, or the $V_\lambda 2$ gene segment can be joined with the $J_\lambda 2$ gene segment. In human or mouse κ light-chain DNA, any one of the V_κ gene segments can be joined with any one of the functional J_λ gene segments.

Rearranged κ and λ genes contain the following regions in order from the 5′ to 3′ end: a short leader (L) exon, a noncoding sequence (intron), a joined VJ gene segment, a second intron, and the constant region. Upstream from each leader gene segment is a promoter sequence. The rearranged light-chain sequence is transcribed by RNA polymerase from the L exon through the C segment to the stop signal, generating a light-chain primary RNA transcript (Figure 5-4). The introns in the primary transcript are removed by RNA-processing enzymes, and the resulting light-chain messenger

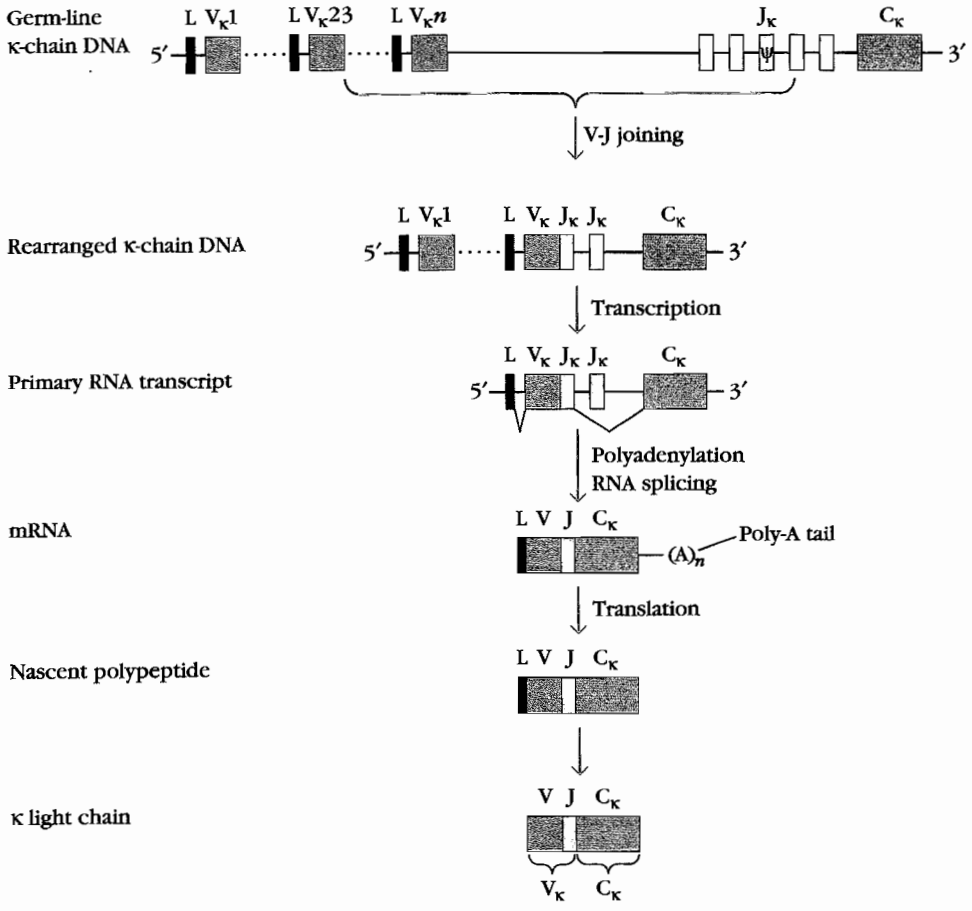


FIGURE 5-4 Kappa light-chain gene rearrangement and RNA processing events required to generate a κ light-chain protein. In this example, rearrangement joins $V_\kappa 23$ and $J_\kappa 4$.

RNA then exits from the nucleus. The light-chain mRNA binds to ribosomes and is translated into the light-chain protein. The leader sequence at the amino terminus pulls the growing polypeptide chain into the lumen of the rough endoplasmic reticulum and is then cleaved, so it is not present in the finished light-chain protein product.

Heavy-Chain DNA Undergoes V-D-J Rearrangements

Generation of a functional immunoglobulin heavy-chain gene requires two separate rearrangement events within the variable region. As illustrated in Figure 5-5, a D_H gene segment first joins to a J_H segment; the resulting D_HJ_H segment then moves next to and joins a V_H segment to generate a $V_HD_HJ_H$ unit that encodes the entire variable region. In heavy-chain DNA, variable-region rearrangement produces a rearranged gene consisting of the following sequences, starting from the 5′ end: a short L exon, an intron, a joined VDJ segment, another intron, and a series of C gene segments. As with the light-chain genes, a promoter sequence is located a short distance upstream from each heavy-chain leader sequence.

Once heavy-chain gene rearrangement is accomplished, RNA polymerase can bind to the promoter sequence and transcribe the entire heavy-chain gene, including the introns. Initially, both C_μ and C_δ gene segments are transcribed. Differential polyadenylation and RNA splicing remove the introns and process the primary transcript to generate mRNA including either the C_μ or the C_δ transcript. These two mRNAs are then translated, and the leader peptide of the resulting nascent polypeptide is cleaved, generating finished μ and δ chains. The production of two different heavy-chain mRNAs allows a mature, immunocompetent B cell to express both IgM and IgD with identical antigenic specificity on its surface.

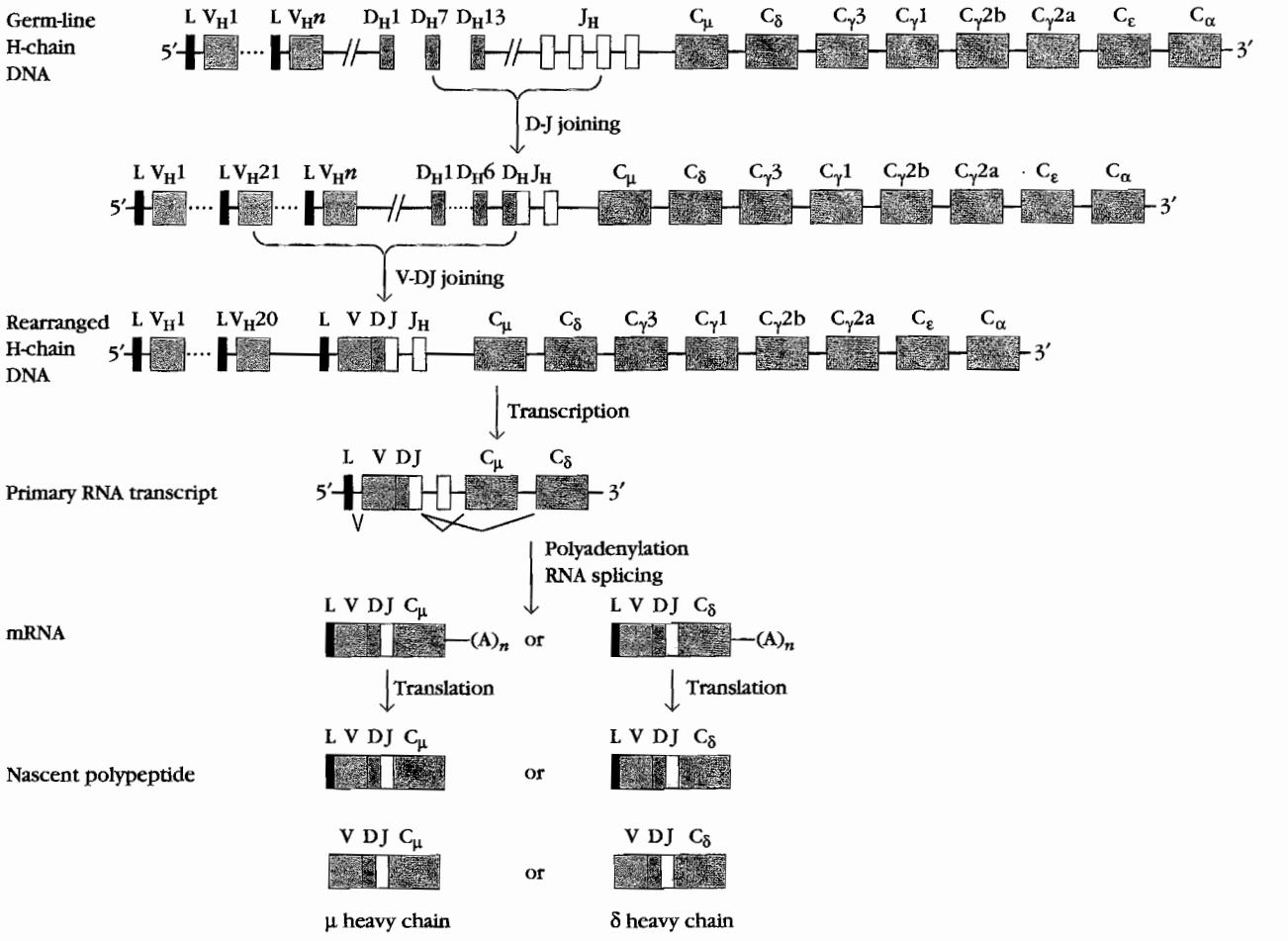


FIGURE 5-5 Heavy-chain gene rearrangement and RNA processing events required to generate finished μ or δ heavy-chain protein. Two DNA joinings are necessary to generate a functional heavy-chain gene: a D_H to J_H joining and a V_H to D_HJ_H joining. In this example, V_H21, D_H7, and J_H3 are joined. Expression of functional heavy-chain genes, although generally similar to expression of light-chain genes, involves differential RNA processing, which generates several different products, including μ or δ heavy chains. Each C gene is drawn as a single coding sequence; in reality, each is organized as a series of exons and introns.

Mechanism of Variable-Region DNA Rearrangements

Now that we've seen the results of variable-region gene rearrangements, let's examine in detail how this process occurs during maturation of B cells.

Recombination Signal Sequences Direct Recombination

The discovery of two closely related conserved sequences in variable-region germ-line DNA paved the way to fuller understanding of the mechanism of gene rearrangements. DNA sequencing studies revealed the presence of unique **recombination signal sequences (RSSs)** flanking each germ-line V, D, and J gene segment. One RSS is located 3′ to each V gene segment, 5′ to each J gene segment, and on both sides of each D gene segment. These sequences function as signals for the recombination process that rearranges the genes. Each RSS contains a conserved palindromic heptamer and a conserved AT-rich nonamer sequence separated by an intervening sequence of 12 or 23 base pairs (Figure 5-6a). The intervening 12- and 23-bp sequences correspond, respectively, to one and two turns of the DNA helix; for this reason the sequences are called **one-turn recombination signal sequences** and **two-turn signal sequences.**

The V_κ signal sequence has a one-turn spacer, and the J_κ signal sequence has a two-turn spacer. In λ light-chain DNA, this order is reversed; that is, the V_λ signal sequence has a two-turn spacer, and the J_λ signal sequence has a one-turn spacer. In heavy-chain DNA, the signal sequences of the V_H and J_H gene segments have two-turn spacers, the signals on either side of the D_H gene segment have one-turn spacers (Figure 5-6b). Signal sequences having a one-turn spacer can join only with sequences having a two-turn spacer (the so-called one-turn/two-turn joining rule). This joining rule ensures, for example, that a V_L segment joins only to a J_L segment and not to another V_L segment; the rule likewise ensures that V_H, D_H, and J_H segments join in proper order and that segments of the same type do not join each other.

Gene Segments Are Joined by Recombinases

V-(D)-J recombination, which takes place at the junctions between RSSs and coding sequences, is catalyzed by enzymes collectively called **V(D)J recombinase.**

Identification of the enzymes that catalyze recombination of V, D, and J gene segments began in the late 1980s and is still ongoing. In 1990 David Schatz, Marjorie Oettinger, and David Baltimore first reported the identification of two **recombination-activating genes,** designated ***RAG-1*** and ***RAG-2***, whose encoded proteins act synergistically and are required to mediate V-(D)-J joining. The RAG-1 and RAG-2 proteins and the enzyme **terminal deoxynucleotidyl transferase** (**TdT**) are the only lymphoid-specific gene products that have been shown to be involved in V-(D)-J rearrangement.

The recombination of variable-region gene segments consists of the following steps, catalyzed by a system of recombinase enzymes (Figure 5-7):

- Recognition of recombination signal sequences (RSSs) by recombinase enzymes, followed by synapsis in which

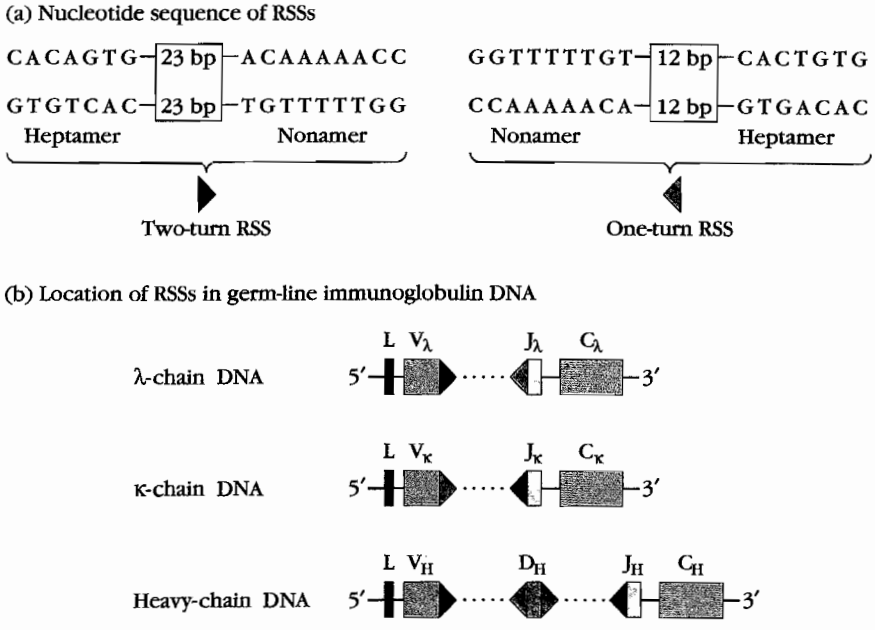


FIGURE 5-6 Two conserved sequences in light-chain and heavy-chain DNA function as recombination signal sequences (RSSs). (a) Both signal sequences consist of a conserved palindromic heptamer and conserved AT-rich nonamer; these are separated by nonconserved spacers of 12 or 23 base pairs. (b) The two types of RSS—designated one-turn RSS and two-turn RSS—have characteristic locations within λ-chain, κ-chain, and heavy-chain germline DNA. During DNA rearrangement, gene segments adjacent to the one-turn RSS can join only with segments adjacent to the two-turn RSS.

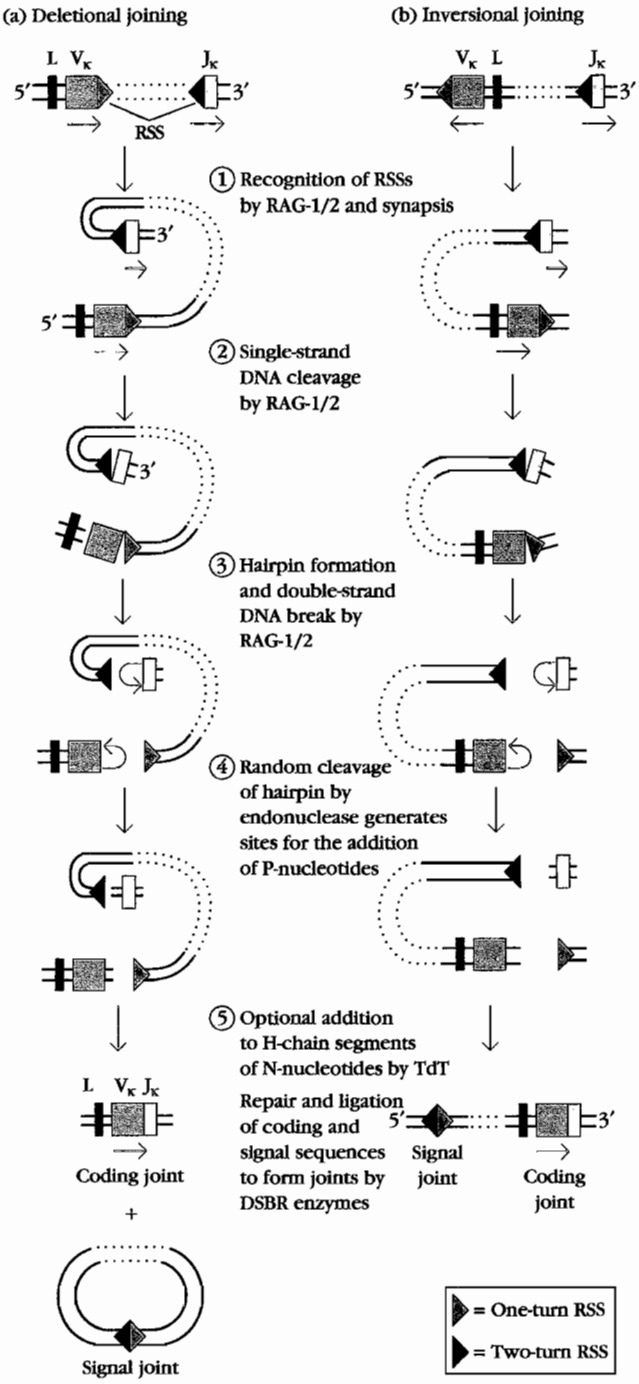


FIGURE 5-7 Model depicting the general process of recombination of immunoglobulin gene segments is illustrated with V_κ and J_κ. (a) Deletional joining occurs when the gene segments to be joined have the same transcriptional orientation (indicated by horizontal blue arrows). This process yields two products: a rearranged VJ unit that includes the coding joint, and a circular excision product consisting of the recombination signal sequences (RSSs), signal joint, and intervening DNA. (b) Inversional joining occurs when the gene segments have opposite transcriptional orientations. In this case, the RSSs, signal joint, and intervening DNA are retained, and the orientation of one of the joined segments is inverted. In both types of recombination, a few nucleotides may be deleted from or added to the cut ends of the coding sequences before they are rejoined.

two signal sequences and the adjacent coding sequences (gene segments) are brought into proximity

- Cleavage of one strand of DNA by RAG-1 and RAG-2 at the junctures of the signal sequences and coding sequences
- A reaction catalyzed by RAG-1 and RAG-2 in which the free 3′-OH group on the cut DNA strand attacks the phosphodiester bond linking the opposite strand to the signal sequence, simultaneously producing a hairpin structure at the cut end of the coding sequence and a flush, 5′-phosphorylated, double-strand break at the signal sequence
- Cutting of the hairpin to generate sites for the addition of **P-region nucleotides,** followed by the trimming of a few nucleotides from the coding sequence by a single-strand endonuclease
- Addition of up to 15 nucleotides, called **N-region nucleotides,** at the cut ends of the V, D, and J coding sequences of the heavy chain by the enzyme terminal deoxynucleotidyl transferase
- Repair and ligation to join the coding sequences and to join the signal sequences, catalyzed by normal double-strand break repair (DSBR) enzymes

Recombination results in the formation of a **coding joint,** falling between the coding sequences, and a **signal joint,** between the RSSs. The transcriptional orientation of the gene segments to be joined determines the fate of the signal joint and intervening DNA. When the two gene segments are in the same transcriptional orientation, joining results in deletion of the signal joint and intervening DNA as a circular excision product (Figure 5-8). Less frequently, the two gene segments have opposite orientations. In this case joining occurs by inversion of the DNA, resulting in the retention of

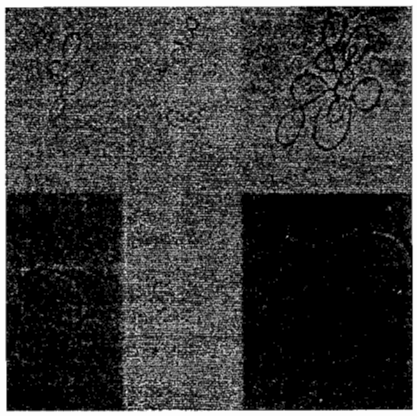


FIGURE 5-8 Circular DNA isolated from thymocytes in which the DNA encoding the chains of the T-cell receptor (TCR) undergoes rearrangement in a process like that involving the immunoglobulin genes. Isolation of this circular excision product is direct evidence for the mechanism of deletional joining shown in Figure 5-7. *[From K. Okazaki et al., 1987,* Cell ***49:****477.]*

both the coding joint and the signal joint (and intervening DNA) on the chromosome. In the human κ locus, about half of the V_κ gene segments are inverted with respect to J_κ and their joining is thus by inversion.

Ig-Gene Rearrangements May Be Productive or Nonproductive

One of the striking features of gene-segment recombination is the diversity of the coding joints that are formed between any two gene segments. Although the double-strand DNA breaks that initiate V-(D)-J rearrangements are introduced precisely at the junctions of signal sequences and coding sequences, the subsequent joining of the coding sequences is imprecise. Junctional diversity at the V-J and V-D-J coding joints is generated by a number of mechanisms: variation in cutting of the hairpin to generate P-nucleotides, variation in trimming of the coding sequences, variation in N-nucleotide addition, and flexibility in joining the coding sequences. The introduction of randomness in the joining process helps generate antibody diversity by contributing to the hypervariability of the antigen-binding site. (This phenomenon is covered in more detail below in the section on generation of antibody diversity.)

Another consequence of imprecise joining is that gene segments may be joined out of phase, so that the triplet reading frame for translation is not preserved. In such a **nonproductive rearrangement,** the resulting VJ or VDJ unit is likely to contain numerous stop codons, which interrupt translation (Figure 5-9). When gene segments are joined in phase, the reading frame is maintained. In such a **productive rearrangement,** the resulting VJ or VDJ unit can be translated in its entirety, yielding a complete antibody.

If one allele rearranges nonproductively, a B cell may still be able to rearrange the other allele productively. If an in-phase rearranged heavy-chain and light-chain gene are not produced, the B cell dies by apoptosis. It is estimated that only one in three attempts at V_L-J_L joining, and one in three attempts at V_H-D_HJ_H joining, are productive. As a result, less than 1/9 (11%) of the early-stage pre-B cells in the bone marrow progress to maturity and leave the bone marrow as mature immunocompetent B cells.

Allelic Exclusion Ensures a Single Antigenic Specificity

B cells, like all somatic cells, are diploid and contain both maternal and paternal chromosomes. Even though a B cell is

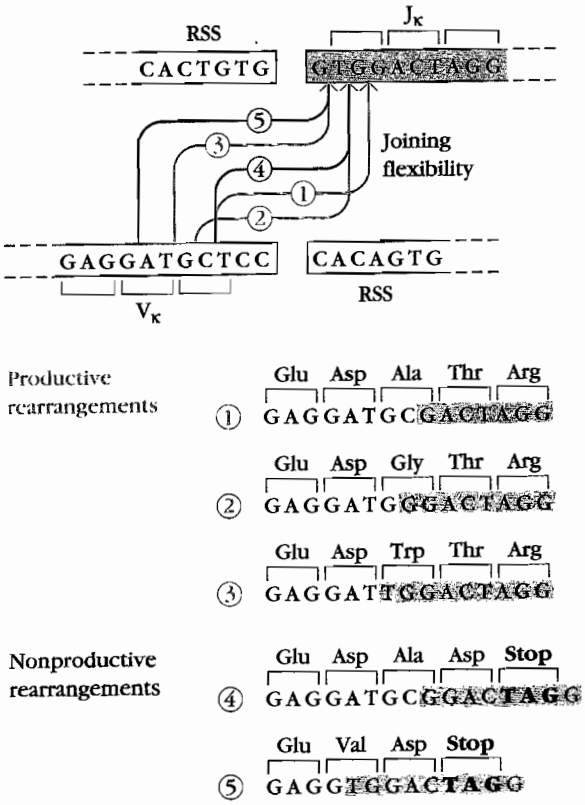


FIGURE 5-9 Junctional flexibility in the joining of immunoglobulin gene segments is illustrated with V_κ and J_κ. In-phase joining (arrows 1, 2, and 3) generates a productive rearrangement, which can be translated into protein. Out-of-phase joining (arrows 4 and 5) leads to a nonproductive rearrangement that contains stop codons and is not translated into protein.

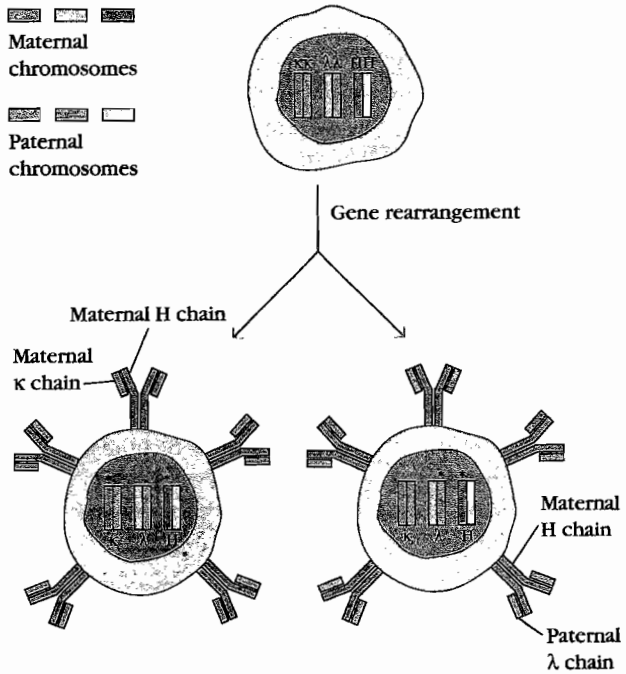


FIGURE 5-10 Because of allelic exclusion, the immunoglobulin heavy- and light-chain genes of only one parental chromosome are expressed per cell. This process ensures that B cells possess a single antigenic specificity. The allele selected for rearrangement is chosen randomly. Thus the expressed immunoglobulin may contain one maternal and one paternal chain or both chains may derive from only one parent. Only B cells and T cells exhibit allelic exclusion. Asterisks (*) indicate the expressed alleles.

diploid, it expresses the rearranged heavy-chain genes from only one chromosome and the rearranged light-chain genes from only one chromosome. The process by which this is accomplished, called **allelic exclusion,** ensures that functional B cells never contain more than one $V_HD_HJ_H$ and one V_LJ_L unit (Figure 5-10). This is, of course, essential for the antigenic specificity of the B cell, because the expression of both alleles would render the B cell multispecific. The phenomenon of allelic exclusion suggests that once a productive V_H-D_H-J_H rearrangement and a productive V_L-J_L rearrangement have occurred, the recombination machinery is turned off, so that the heavy- and light-chain genes on the homologous chromosomes are not expressed.

G. D. Yancopoulos and F. W. Alt have proposed a model to account for allelic exclusion (Figure 5-11). They suggest that once a productive rearrangement is attained, its encoded protein is expressed and the presence of this protein acts as a signal to prevent further gene rearrangement. According to their model, the presence of μ heavy chains signals the maturing B cell to turn off rearrangement of the other heavy-chain allele and to turn on rearrangement of the κ light-chain genes. If a productive κ rearrangement occurs, κ light chains are produced and then pair with μ heavy chains to form a complete antibody molecule. The presence of this antibody then turns off further light-chain rearrangement. If κ rearrangement is nonproductive for both κ alleles, rearrangement of the λ-chain genes begins. If neither λ allele rearranges productively, the B cell presumably ceases to mature and soon dies by apoptosis.

Two studies with transgenic mice have supported the hypothesis that the protein products encoded by rearranged heavy- and light-chain genes regulate rearrangement of the remaining alleles. In one study, transgenic mice carrying a rearranged μ heavy-chain transgene were prepared. The μ transgene product was expressed by a large percentage of the B cells, and rearrangement of the endogenous immunoglobulin heavy-chain genes was blocked. Similarly, cells from a transgenic mouse carrying a κ light-chain transgene did not

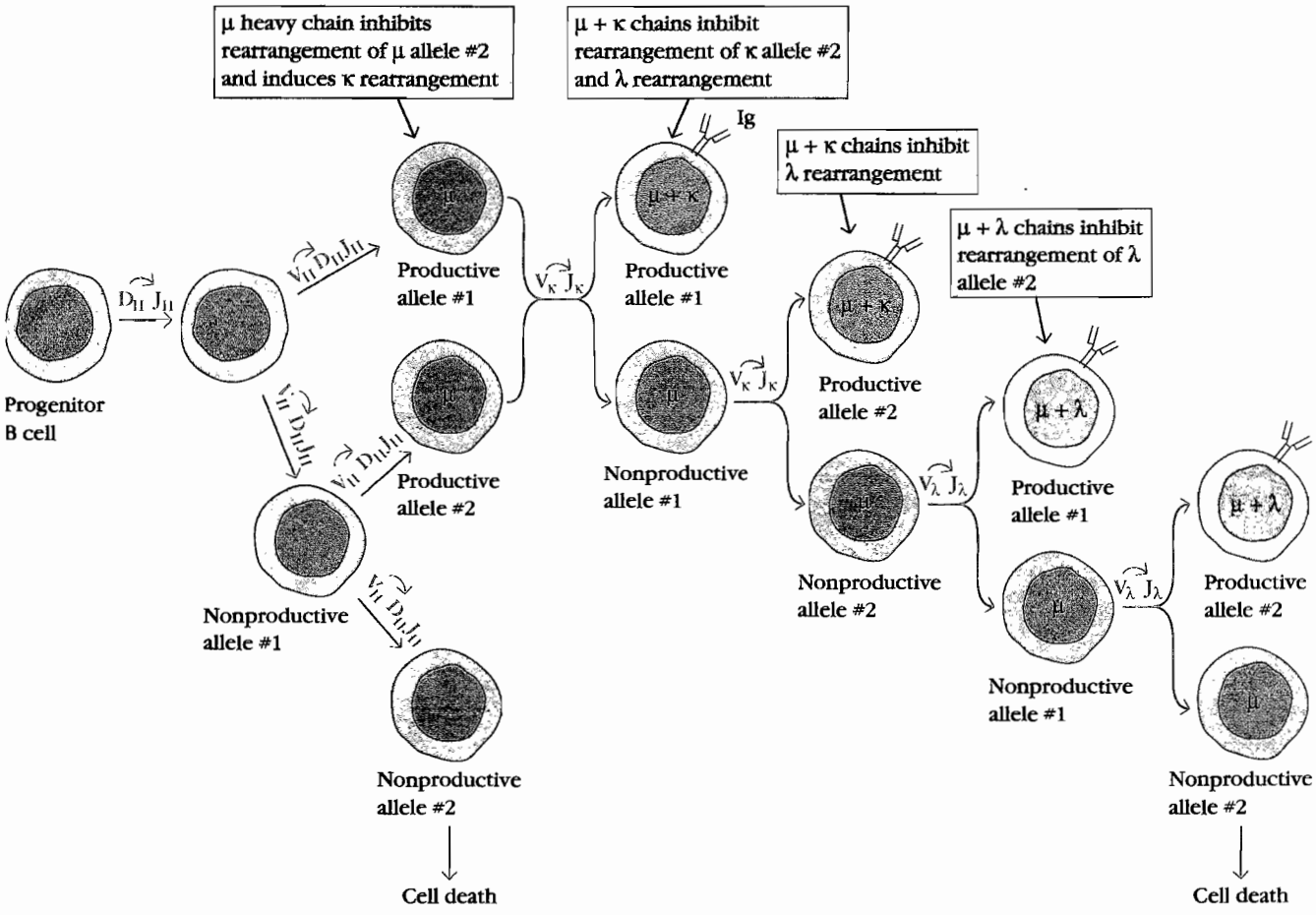


FIGURE 5-11 Model to account for allelic exclusion. Heavy-chain genes rearrange first, and once a productive heavy-chain gene rearrangement occurs, the μ protein product prevents rearrangement of the other heavy-chain allele and initiates light-chain gene rearrangement. In the mouse, rearrangement of κ light-chain genes precedes rearrangement of the λ genes, as shown here. In humans, either κ or λ rearrangement can proceed once a productive heavy-chain rearrangement has occurred. Formation of a complete immunoglobulin inhibits further light-chain gene rearrangement. If a nonproductive rearrangement occurs for one allele, then the cell attempts rearrangement of the other allele. *[Adapted from G. D. Yancopoulos and F. W. Alt, 1986,* Annu. Rev. Immunol. ***4:****339.]*

rearrange the endogenous κ-chain genes when the κ transgene was expressed and was associated with a heavy chain to form complete immunoglobulin. These studies suggest that expression of the heavy- and light-chain proteins may indeed prevent gene rearrangement of the remaining alleles and thus account for allelic exclusion.

Generation of Antibody Diversity

As the organization of the immunoglobulin genes was deciphered, the sources of the vast diversity in the variable region began to emerge. The germ-line theory, mentioned earlier, argued that the entire variable-region repertoire is encoded in the germ line of the organism and is transmitted from parent to offspring through the germ cells (egg and sperm). The somatic-variation theory held that the germ line contains a limited number of variable genes, which are diversified in the somatic cells by mutational or recombinational events during development of the immune system. With the cloning and sequencing of the immunoglobulin genes, both models were partly vindicated.

To date, seven means of antibody diversification have been identified in mice and humans:

- Multiple germ-line gene segments
- Combinatorial V-(D)-J joining
- Junctional flexibility
- P-region nucleotide addition (P-addition)
- N-region nucleotide addition (N-addition)
- Somatic hypermutation
- Combinatorial association of light and heavy chains

Although the exact contribution of each of these avenues of diversification to total antibody diversity is not known, they each contribute significantly to the immense number of distinct antibodies that the mammalian immune system is capable of generating.

There Are Numerous Germ-Line V, D, and J Gene Segments

An inventory of functional V, D, and J gene segments in the germ-line DNA of one human reveals 51 V_H, 25 D, 6 J_H, 40 V_κ, 5 J_κ, 31 V_λ, and 4 J_λ gene segments. In addition to these functional segments, there are many pseudogenes. It should be borne in mind that these numbers were largely derived from a landmark study that sequenced the DNA of the immunoglobulin loci of a single individual. The immunoglobulin loci of other individuals might contain slightly different numbers of particular types of gene segments.

In the mouse, although the numbers are known with less precision than in the human, there appear to be about 85 V_κ gene segments and 134 V_H gene segments, 4 functional J_H, 4 functional J_κ, 3 functional J_λ, and an estimated 13 D_H gene segments, but only three V_λ gene segments. Although the number of germ-line genes found in either humans or mice is far fewer than predicted by early proponents of the germ-line model, multiple germ-line V, D, and J gene segments clearly do contribute to the diversity of the antigen-binding sites in antibodies.

Combinatorial V-J and V-D-J Joining Generates Diversity

The contribution of multiple germ-line gene segments to antibody diversity is magnified by the random rearrangement of these segments in somatic cells. It is possible to calculate how much diversity can be achieved by gene rearrangments (Table 5-2). In humans, the ability of any of the 51 V_H gene segments to combine with any of the 27 D_H segments and any of the 6 J_H segments allows a considerable amount of heavy-chain gene diversity to be generated ($51 \times 27 \times 6 = 8262$ possible combinations). Similarly, 40 V_κ gene segments randomly combining with 5 J_κ segments has the potential of generating 200 possible combinations at the κ locus, while 30 V_λ and 4 J_λ gene segments allow up to 120 possible combinations at the human λ locus. It is important to realize that these are minimal calculations of potential diversity. Junctional flexibility and P- and N-nucleotide addition, as mentioned above, and, especially, somatic hypermutation, which will be described shortly, together make an enormous contribution to antibody diversity. Although it is not possible to make an exact calculation of their contribution, most workers in this field agree that they raise the potential for antibody combining-site diversity in humans to well over 10^{10}. This does not mean that, at any given time, a single individual has a repertoire of 10^{10} different antibody combining sites. These very large numbers describe the set of possible variations, of which any individual carries a subset that is smaller by several orders of magnitude.

Junctional Flexibility Adds Diversity

The enormous diversity generated by means of V, D, and J combinations is further augmented by a phenomenon called **junctional flexibility.** As described above, recombination involves both the joining of recombination signal sequences to form a signal joint and the joining of coding sequences to form a coding joint (see Figure 5-7). Although the signal sequences are always joined precisely, joining of the coding sequences is often imprecise. In one study, for example, joining of the $V_\kappa 21$ and $J_\kappa 1$ coding sequences was analyzed in several pre-B cell lines. Sequence analysis of the signal and coding joints revealed the contrast in junctional precision (Figure 5-12).

As illustrated previously, junctional flexibility leads to many nonproductive rearrangements, but it also generates productive combinations that encode alternative amino acids at each coding joint (see Figure 5-9), thereby increasing antibody diversity. The amino acid sequence variation gener-

TABLE 5-2 Combinatorial antibody diversity in humans and mice

Multiple germ-line segments	Heavy chain	LIGHT CHAINS κ	LIGHT CHAINS λ
	ESTIMATED NUMBER OF SEGMENTS IN HUMANS*		
V	51	40	30
D	27	0	0
J	6	5	4
Combinatorial V-D-J and V-J joining (possible number of combinations)	51 × 27 × 6 = 8262	40 × 5 = 200	30 × 4 = 120
Possible combinatorial associations of heavy and light chains†		8262 × (200 × 120) = 2.64 × 10^6	
	ESTIMATED NUMBER OF SEGMENTS IN MICE*		
V	134	85	2
D	13	0	0
J	4	4	3
Combinatorial V-D-J and V-J joining (possible number of combinations)	134 × 13 × 4 = 6968	85 × 4 = 340	2 × 3 = 6
Possible combinatorial associations of heavy and light chains†		6968 × (340 + 6) = 2.41 × 10^6	

*These numbers have been determined from studies of single subjects; slight differences may be seen among different individuals. Also, in the human case, only the functional gene segments have been listed. The genome contains additional segments that are incapable of rearrangement or contain stop codons or both. In the mouse case, the figures contained in the table are only best estimates, because the locus has not been completely sequenced.

†Because of the diversity contributed by junctional flexibility, P-region nucleotide addition, N-region nucleotide addition, and somatic mutation, the actual potential exceeds these estimates by several orders of magnitude.

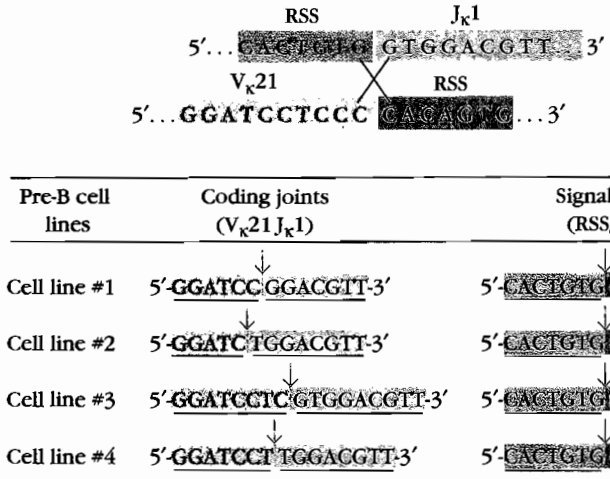


FIGURE 5-12 Experimental evidence for junctional flexibility in immunoglobulin-gene rearrangement. The nucleotide sequences flanking the coding joints between $V_κ21$ and $J_κ1$ and the corresponding signal joint sequences were determined in four pre-B cell lines. The sequence constancy in the signal joints contrasts with the sequence variability in the coding joints. Pink and yellow shading indicate nucleotides derived from $V_κ21$ and $J_κ1$, respectively, and purple and orange shading indicate nucleotides from the two RSSs.

ated by junctional flexibility in the coding joints has been shown to fall within the third hypervariable region (CDR3) in immunoglobulin heavy-chain and light-chain DNA (Table 5-3). Since CDR3 often makes a major contribution to antigen binding by the antibody molecule, amino acid changes generated by junctional flexibility are important in the generation of antibody diversity.

P-Addition Adds Diversity at Palindromic Sequences

As described earlier, after the initial single-strand DNA cleavage at the junction of a variable-region gene segment and attached signal sequence, the nucleotides at the end of the coding sequence turn back to form a hairpin structure (see Figure 5-7). This hairpin is later cleaved by an endonuclease. This second cleavage sometimes occurs at a position that leaves a short single strand at the end of the coding sequence. The subsequent addition of complementary nucleotides to this strand (**P-addition**) by repair enzymes generates a palindromic sequence in the coding joint, and so these nucleotides are called **P-nucleotides** (Figure 5-13a). Variation in the position at which the hairpin is cut thus leads to variation in the sequence of the coding joint.

N-Addition Adds Considerable Diversity by Addition of Nucleotides

Variable-region coding joints in rearranged heavy-chain genes have been shown to contain short amino acid sequences that are not encoded by the germ-line V, D, or J gene segments. These amino acids are encoded by nucleotides added during the D-J and V to D-J joining process by a terminal deoxynucleotidyl transferase (TdT) catalyzed reaction (Figure 5-13b). Evidence that TdT is responsible for the addition of these **N-nucleotides** has come from transfection studies in fibroblasts. When fibroblasts were transfected with the *RAG-1* and *RAG-2* genes, V-D-J rearrangement occurred but no N-nucleotides were present in the coding joints. However, when the fibroblasts were also transfected with the gene encoding TdT, then V-D-J rearrangement was accompanied by addition of N-nucleotides at the coding joints.

Up to 15 N-nucleotides can be added to both the D_H-J_H and V_H-D_HJ_H joints. Thus, a complete heavy-chain variable region is encoded by a $V_HND_HNJ_H$ unit. The additional heavy-chain diversity generated by N-region nucleotide addition is quite large because N regions appear to consist of wholly random sequences. Since this diversity occurs at V-D-J coding joints, it is localized in CDR3 of the heavy-chain genes.

TABLE 5-3 Sources of sequence variation in complementarity-determining regions of immunoglobulin heavy- and light-chain genes

Source of variation	CDR1	CDR2	CDR3
Sequence encoded by:	V segment	V segment	V_L-J_L junction; V_H-D_H-J_H junctions
Junctional flexibility	–	–	+
P-nucleotide addition	–	–	+
N-nucleotide addition*	–	–	+
Somatic hypermutation	+	+	+

*N-nucleotide addition occurs only in heavy-chain DNA.

Somatic Hypermutation Adds Diversity in Already-Rearranged Gene Segments

All the antibody diversity described so far stems from mechanisms that operate during formation of specific variable regions by gene rearrangement. Additional antibody diversity is generated in rearranged variable-region gene units by a process called **somatic hypermutation.** As a result of somatic hypermutation, individual nucleotides in VJ or VDJ units are replaced with alternatives, thus potentially altering the specificity of the encoded immunoglobulins.

Normally, somatic hypermutation occurs only within germinal centers (see Chapter 11), structures that form in secondary lymphoid organs within a week or so of immunization with an antigen that activates a T-cell-dependent B-cell response. Somatic hypermutation is targeted to rearranged V-regions located within a DNA sequence containing about 1500 nucleotides, which includes the whole of the VJ or VDJ segment. Somatic hypermutation occurs at a frequency approaching 10^{-3} per base pair per generation. This rate is at least a hundred thousand-fold higher (hence the name *hypermutation*) than the spontaneous mutation rate, about 10^{-8}/bp/generation, in other genes. Since the combined length of the H-chain and L-chain variable-region genes is about 600 bp, one expects that somatic hypermutation will introduce at least one mutation per every two cell divisions in the pair of V_H and V_L genes that encode an antibody.

The mechanism of somatic hypermutation has not yet been determined. Most of the mutations are nucleotide substitutions rather than deletions or insertions. Somatic hypermutation introduces these substitutions in a largely, but not completely, random fashion. Recent evidence suggests that certain nucleotide motifs and palindromic sequences within V_H and V_L may be especially susceptible to somatic hypermutation.

Somatic hypermutations occur throughout the VJ or VDJ segment, but in mature B cells they are clustered within the CDRs of the V_H and V_L sequences, where they are most likely to influence the overall affinity for antigen. Following exposure to antigen, those B cells with higher-affinity receptors will be preferentially selected for survival. The result of this

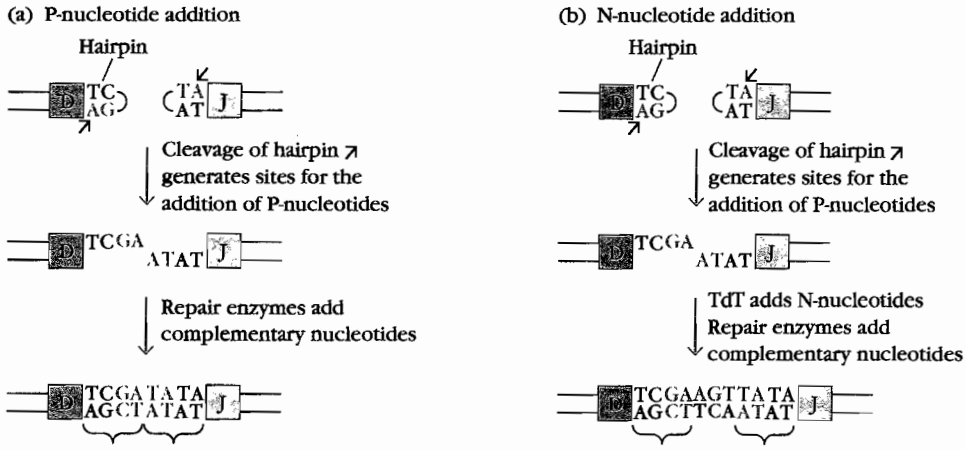


FIGURE 5-13 P-nucleotide and N-nucleotide addition during joining. (a) If cleavage of the hairpin intermediate yields a double-stranded end on the coding sequence, then P-nucleotide addition does not occur. In many cases, however, cleavage yields a single-stranded end. During subsequent repair, complementary nucleotides are added, called P-nucleotides, to produce palindromic sequences (indicated by brackets). In this example, four extra base pairs (blue) are present in the coding joint as the result of P-nucleotide addition. (b) Besides P-nucleotide addition, addition of random N-nucleotides (light red) by a terminal deoxynucleotidyl transferase (TdT) can occur during joining of heavy-chain coding sequences.

differential selection is an increase in the antigen affinity of a population of B cells. The overall process, called **affinity maturation,** takes place within germinal centers, and is described more fully in Chapter 11.

Claudia Berek and Cesar Milstein obtained experimental evidence demonstrating somatic hypermutation during the course of an immune response to a hapten-carrier conjugate. These researchers were able to sequence mRNA that encoded antibodies raised against a hapten in response to primary, secondary, or tertiary immunization (first, second, or third exposure) with a hapten-carrier conjugate. The hapten they chose was 2-phenyl-5-oxazolone (phOx), coupled to a protein carrier. They chose this hapten because it had previously been shown that the majority of antibodies it induced were encoded by a single germ-line V_H and V_κ gene segment. Berek and Milstein immunized mice with the phOx-carrier conjugate and then used the mouse spleen cells to prepare hybridomas secreting monoclonal antibodies specific for the phOx hapten. The mRNA sequence for the H chain and κ light chain of each hybridoma was then determined to identify deviations from the germ-line sequences.

The results of this experiment are depicted in Figure 5-14. Of the 12 hybridomas obtained from mice seven days after a primary immunization, all used a particular V_H, the V_H Ox-1 gene segment, and all but one used the same V_L gene segment, V_κ Ox-1. Moreover, only a few mutations from the germ-line sequence were present in these hybridomas. By day 14 after primary immunization, analysis of eight hybridomas revealed that six continued to use the germ-line V_H Ox-1 gene segment and all continued to use the V_κ Ox-1 gene segment. Now, however, all of these hybridomas included one or more mutations from the germ-line sequence. Hybridomas analyzed from the secondary and tertiary responses showed a larger percentage utilizing germ-line V_H gene segments other than the V_H Ox-1 gene. In those hybridoma clones that utilized the V_H Ox-1 and V_κ Ox-1 gene segments, most of the mutations were clustered in the CDR1 and CDR2 hypervariable regions. The number of mutations in the anti-phOx hybridomas progressively increased following primary, secondary, and tertiary immunizations, as did the overall affinity of the antibodies for phOx (see Figure 5-14).

A Final Source of Diversity Is Combinatorial Association of Heavy and Light Chains

In humans, there is the potential to generate 8262 heavy-chain genes and 320 light-chain genes as a result of variable-region gene rearrangements. Assuming that any one of the possible heavy-chain and light-chain genes can occur randomly in the same cell, the potential number of heavy- and light-chain combinations is 2,644,240. This number is probably higher than the amount of combinatorial diversity actually generated in an individual, because it is not likely that all V_H and V_L will pair with each other. Furthermore, the recombination process is not completely random; not all V_H, D, or V_L gene segments are used at the same frequency. Some are used often, others only occasionally, and still others almost never.

Although the number of different antibody combining sites the immune system can generate is difficult to calculate with precision, we know that it is quite high. Because the very large number of new sequences created by junctional

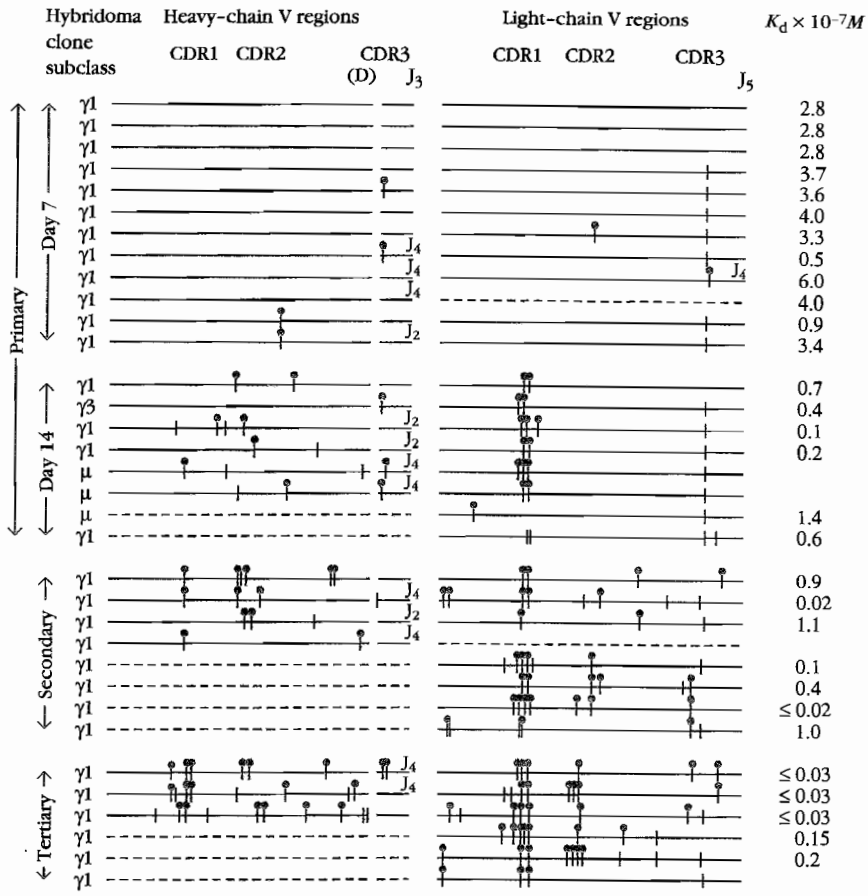


FIGURE 5.14 Experimental evidence for somatic mutation in variable regions of immunoglobulin genes. The diagram compares the mRNA sequences of heavy chains and of light chains from hybridomas specific for the phOx hapten. The horizontal solid lines represent the germ-line V_H and V_κ Ox-1 sequences; dashed lines represent sequences derived from other germ-line genes. Blue shading shows the areas where mutations clustered; the blue circles with vertical lines indicate locations of mutations that encode a different amino acid than the germ-line sequence. These data show that the frequency of mutation (1) increases in the course of the primary response (day 7 vs. day 14) and (2) is higher after secondary and tertiary immunizations than after primary immunization. Moreover, the dissociation constant (K_d) of the anti-phOx antibodies decreases during the transition from the primary to tertiary response, indicating an increase in the overall affinity of the antibody. Note also that most of the mutations are clustered within CDR1 and CDR2 of both the heavy and the light chains. *[Adapted from C. Berek and C. Milstein, 1987,* Immunol. Rev. ***96:****23.]*

flexibility, P-nucleotide addition, and N-nucleotide addition are within the third CDR, they are positioned to influence the structure of the antibody binding site. In addition to these sources of antibody diversity, the phenomenon of somatic hypermutation contributes enormously to the repertoire after antigen stimulation.

Class Switching among Constant-Region Genes

After antigenic stimulation of a B cell, the heavy-chain DNA can undergo a further rearrangement in which the $V_HD_HJ_H$ unit can combine with any C_H gene segment. The exact mechanism of this process, called **class switching** or **isotype switching,** is unclear, but it involves DNA flanking sequences (called **switch regions**) located 2–3 kb upstream from each C_H segment (except C_δ). These switch regions, though rather large (2 to 10 kb), are composed of multiple copies of short repeats (GAGCT and TGGGG). One hypothesis is that a protein or system of proteins that constitute the switch recombinase recognize these repeats and upon binding carry out the DNA recombination that results in class switching. Intercellular regulatory proteins known as cytokines act as "switch factors" and play major roles in determining the particular immunoglobulin class that is expressed as a consequence of switching. Interleukin 4 (IL-4),

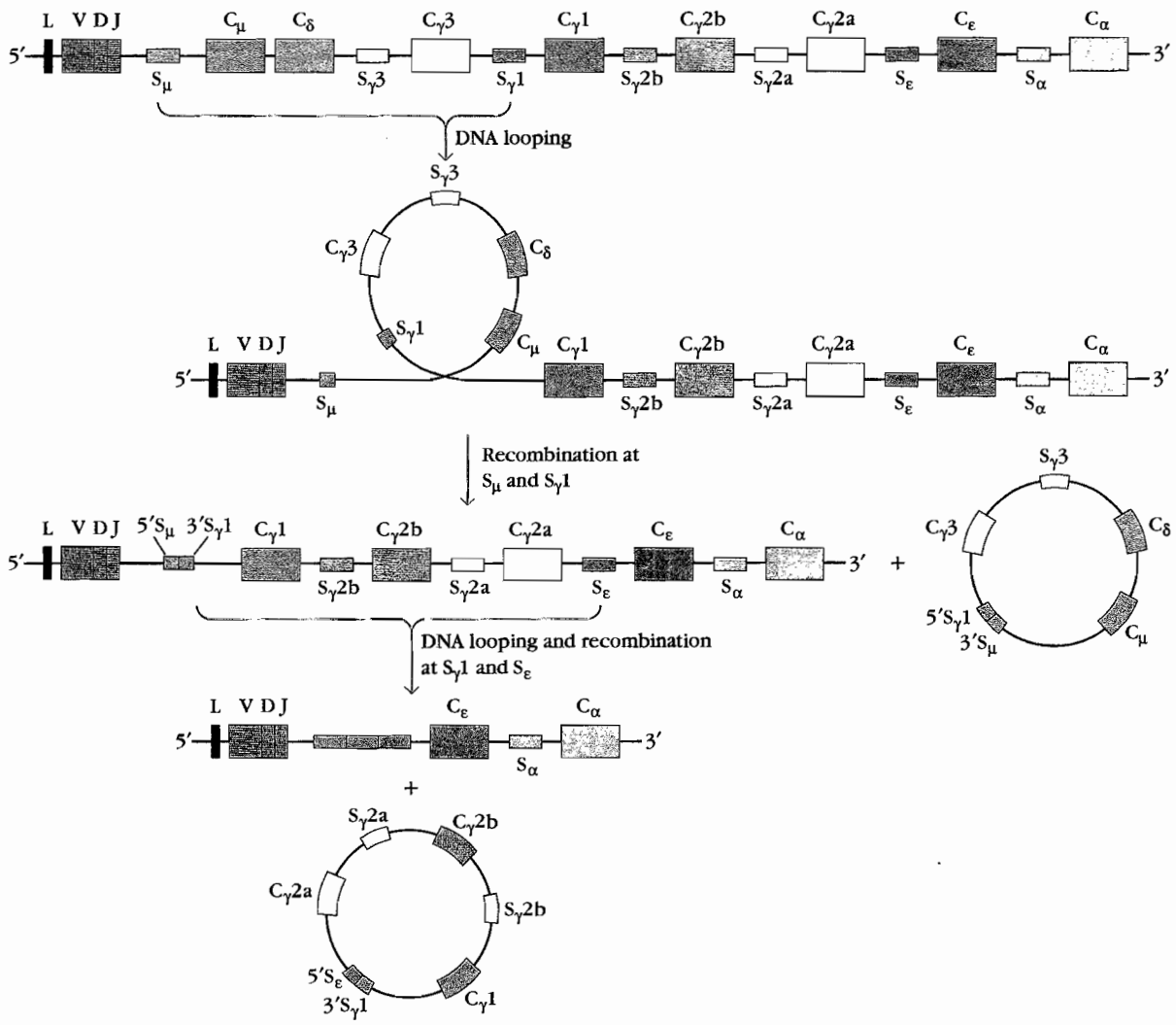


FIGURE 5-15 Proposed mechanism for class switching induced by interleukin 4 in rearranged immunoglobulin heavy-chain genes. A switch site is located upstream from each C_H segment except C_δ. Identification of the indicated circular excision products containing portions of the switch sites suggested that IL-4 induces sequential class switching from C_μ to $C_\gamma1$ to C_ϵ.

for example, induces class switching from C_μ to $C_\gamma1$ or C_ϵ. In some cases, IL-4 has been observed to induce class switching in a successive manner: first from C_μ to $C_\gamma1$ and then from $C_\gamma1$ to C_ϵ (Figure 5-15). Examination of the DNA excision products produced during class switching from C_μ to $C_\gamma1$ showed that a circular excision product containing C_μ together with the 5′ end of the $\gamma1$ switch region ($S_\gamma1$) and the 3′ end of the μ switch region (S_μ) was generated. Furthermore, the switch from $C_\gamma1$ to C_ϵ produced circular excision products containing $C_\gamma1$ together with portions of the μ, γ, and ϵ switch regions. Thus class switching depends upon the interplay of three elements: switch regions, a switch recombinase, and the cytokine signals that dictate the isotype to which the B cell switches. A more complete description of the role of cytokines in class switching appears in Chapter 11.

Expression of Ig Genes

As in the expression of other genes, post-transcriptional processing of immunoglobulin primary transcripts is required to produce functional mRNAs (see Figures 5-4 and 5-5). The primary transcripts produced from rearranged heavy-chain and light-chain genes contain intervening DNA sequences that include noncoding introns and J gene segments not lost during V-(D)-J rearrangement. In addition, as noted earlier, the heavy-chain C-gene

segments are organized as a series of coding exons and noncoding introns. Each exon of a C_H gene segment corresponds to a constant-region domain or a hinge region of the heavy-chain polypeptide. The primary transcript must be processed to remove the intervening DNA sequences, and the remaining exons must be connected by a process called RNA splicing. Short, moderately conserved splice sequences, or splice sites, which are located at the intron-exon boundaries within a primary transcript, signal the positions at which splicing occurs. Processing of the primary transcript in the nucleus removes each of these intervening sequences to yield the final mRNA product. The mRNA is then exported from the nucleus to be translated by ribosomes into complete H or L chains.

Heavy-Chain Primary Transcripts Undergo Differential RNA Processing

Processing of an immunoglobulin heavy-chain primary transcript can yield several different mRNAs, which explains how a single B cell can produce secreted or membrane-bound forms of a particular immunoglobulin and simultaneously express IgM and IgD.

EXPRESSION OF MEMBRANE OR SECRETED IMMUNOGLOBULIN

As explained in Chapter 4, a particular immunoglobulin can exist in either membrane-bound or secreted form. The two forms differ in the amino acid sequence of the heavy-chain carboxyl-terminal domains (C_H3/C_H3 in IgA, IgD, and IgG and C_H4/C_H4 in IgE and IgM). The secreted form has a hydrophilic sequence of about 20 amino acids in the carboxyl-terminal domain; this is replaced in the membrane-bound form with a sequence of about 40 amino acids containing a hydrophilic segment that extends outside the cell, a hydrophobic transmembrane segment, and a short hydrophilic segment at the carboxyl terminus that extends into the cytoplasm (Figure 5-16a). For some time, the existence of these two forms seemed inconsistent with the structure of germline heavy-chain DNA, which had been shown to contain a single C_H gene segment corresponding to each class and subclass.

The resolution of this puzzle came from DNA sequencing of the C_μ gene segment, which consists of four exons ($C_\mu1$, $C_\mu2$, $C_\mu3$, and $C_\mu4$) corresponding to the four domains of the IgM molecule. The $C_\mu4$ exon contains a nucleotide sequence (called S) at its 3′ end that encodes the hydrophilic sequence in the C_H4 domain of secreted IgM. Two additional exons called M1 and M2 are located just 1.8 kb downstream from the 3′ end of the $C_\mu4$ exon. The M1 exon encodes the transmembrane segment, and M2 encodes the cytoplasmic segment of the C_H4 domain in membrane-bound IgM. Later DNA sequencing revealed that all the C_H gene segments have two additional downstream M1 and M2 exons that encode the transmembrane and cytoplasmic segments.

The primary transcript produced by transcription of a rearranged μ heavy-chain gene contains two polyadenylation signal sequences, or poly-A sites, in the C_μ segment. Site 1 is located at the 3′ end of the $C_\mu4$ exon, and site 2 is at the 3′ end of the M2 exon (Figure 5-16b). If cleavage of the primary transcript and addition of the poly-A tail occurs at site 1, the M1 and M2 exons are lost. Excision of the introns and splicing of the remaining exons then produces mRNA encoding the secreted form of the heavy chain. If cleavage and polyadenylation of the primary transcript occurs instead at site 2, then a different pattern of splicing results. In this case, splicing removes the S sequence at the 3′ end of the $C_\mu4$ exon, which encodes the hydrophilic carboxyl-terminal end of the secreted form, and joins the remainder of the $C_\mu4$ exon with the M1 and M2 exons, producing mRNA for the membrane form of the heavy chain.

Thus, differential processing of a common primary transcript determines whether the secreted or membrane form of an immunoglobulin will be produced. As noted previously, mature naive B cells produce only membrane-bound antibody, whereas differentiated plasma cells produce secreted antibodies. It remains to be determined precisely how naive B cells and plasma cells direct RNA processing preferentially toward the production of mRNA encoding one form or the other.

SIMULTANEOUS EXPRESSION OF IgM AND IgD

Differential RNA processing also underlies the simultaneous expression of membrane-bound IgM and IgD by mature B cells. As mentioned already, transcription of rearranged heavy-chain genes in mature B cells produces primary transcripts containing both the C_μ and C_δ gene segments. The C_μ and C_δ, gene segments are close together in the rearranged gene (only about 5 kb apart), and the lack of a switch site between them permits the entire $VDJC_\mu C_\delta$ region to be transcribed into a single primary RNA transcript about 15 kb long, which contains four poly-A sites (Figure 5-17a). Sites 1 and 2 are associated with C_μ, as described in the previous section; sites 3 and 4 are located at similar places in the C_δ gene segment. If the heavy-chain transcript is cleaved and polyadenylated at site 2 after the C_μ exons, then the mRNA will encode the membrane form of the μ heavy chain (Figure 5-17b); if polyadenylation is instead further downstream at site 4, after the C_δ exons, then RNA splicing will remove the intervening C_μ exons and produce mRNA encoding the membrane form of the δ heavy chain (Figure 5-17c).

Since the mature B cell expresses both IgM and IgD on its membrane, both processing pathways must occur simultaneously. Likewise, cleavage and polyadenylation of the primary heavy-chain transcript at poly-A site 1 or 3 in

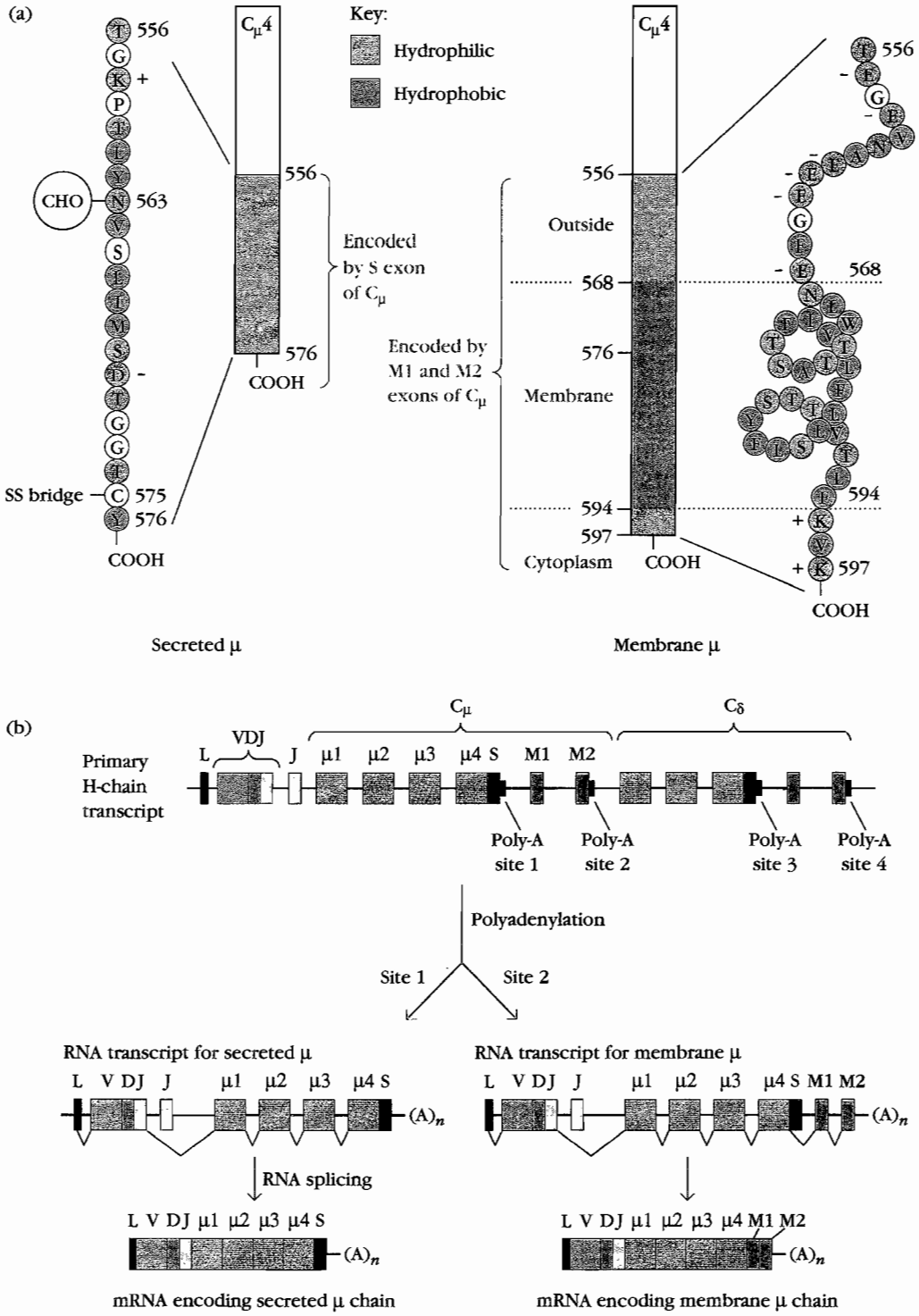


FIGURE 5-16 Expression of secreted and membrane forms of the heavy chain by alternative RNA processing. (a) Amino acid sequence of the carboxyl-terminal end of secreted and membrane μ heavy chains. Residues are indicated by the single-letter amino acid code. Hydrophilic and hydrophobic residues and regions are indicated by purple and orange, respectively, and charged amino acids are indicated with a + or −. The white regions of the sequences are identical in both forms. (b) Structure of the primary transcript of a rearranged heavy-chain gene showing the C_μ exons and poly-A sites. Polyadenylation of the primary transcript at either site 1 or site 2 and subsequent splicing (indicated by V-shaped lines) generates mRNAs encoding either secreted or membrane μ chains.

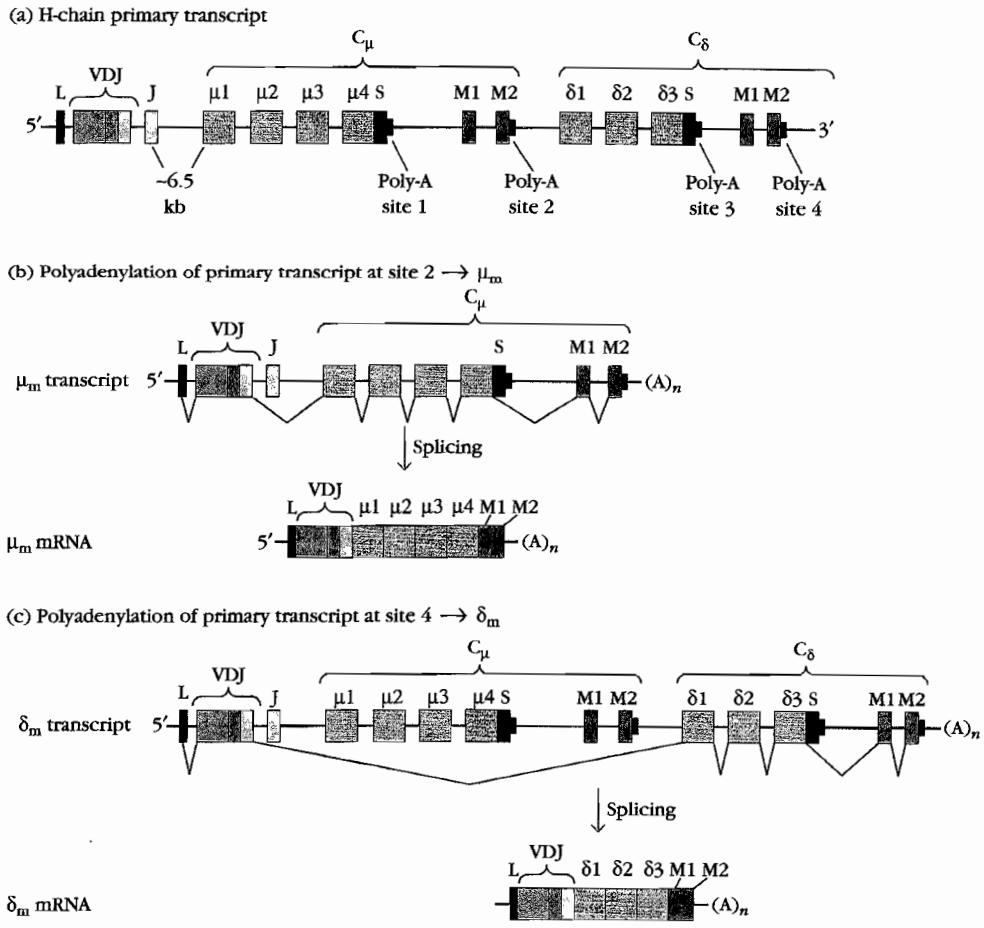


FIGURE 5-17 Expression of membrane forms of μ and δ heavy chains by alternative RNA processing. (a) Structure of rearranged heavy-chain gene showing C_μ and C_δ exons and poly-A sites. (b) Structure of μ_m transcript and μ_m mRNA resulting from polyadenylation at site 2 and splicing. (c) Structure of δ_m transcript and δ_m mRNA resulting from polyadenylation at site 4 and splicing. Both processing pathways can proceed in any given B cell.

plasma cells and subsequent splicing will yield the secreted form of the μ or δ heavy chains, respectively (see Figure 5-16b).

Synthesis, Assembly, and Secretion of Immunoglobulins

Immunoglobulin heavy- and light-chain mRNAs are translated on separate polyribosomes of the rough endoplasmic reticulum (RER). Newly synthesized chains contain an amino-terminal leader sequence, which serves to guide the chains into the lumen of the RER, where the signal sequence is then cleaved. The assembly of light (L) and heavy (H) chains into the disulfide-linked and glycosylated immunoglobulin molecule occurs as the chains pass through the cisternae of the RER. The complete molecules are transported to the Golgi apparatus and then into secretory vesicles, which fuse with the plasma membrane (Figure 5-18).

The order of chain assembly varies among the immunoglobulin classes. In the case of IgM, the H and L chains assemble within the RER to form half-molecules, and then two half-molecules assemble to form the complete molecule. In the case of IgG, two H chains assemble, then an H_2L intermediate is assembled, and finally the complete H_2L_2 molecule is formed. Interchain disulfide bonds are formed, and the polypeptides are glycosylated as they move through the Golgi apparatus.

If the molecule contains the transmembrane sequence of the membrane form, it becomes anchored in the membrane of a secretory vesicle and is inserted into the plasma membrane as the vesicle fuses with the plasma membrane (see Figure 5-18, insert). If the molecule contains the hydrophilic sequence of secreted immunoglobulins, it is transported as a free molecule in a secretory vesicle and is released from the cell when the vesicle fuses with the plasma membrane.

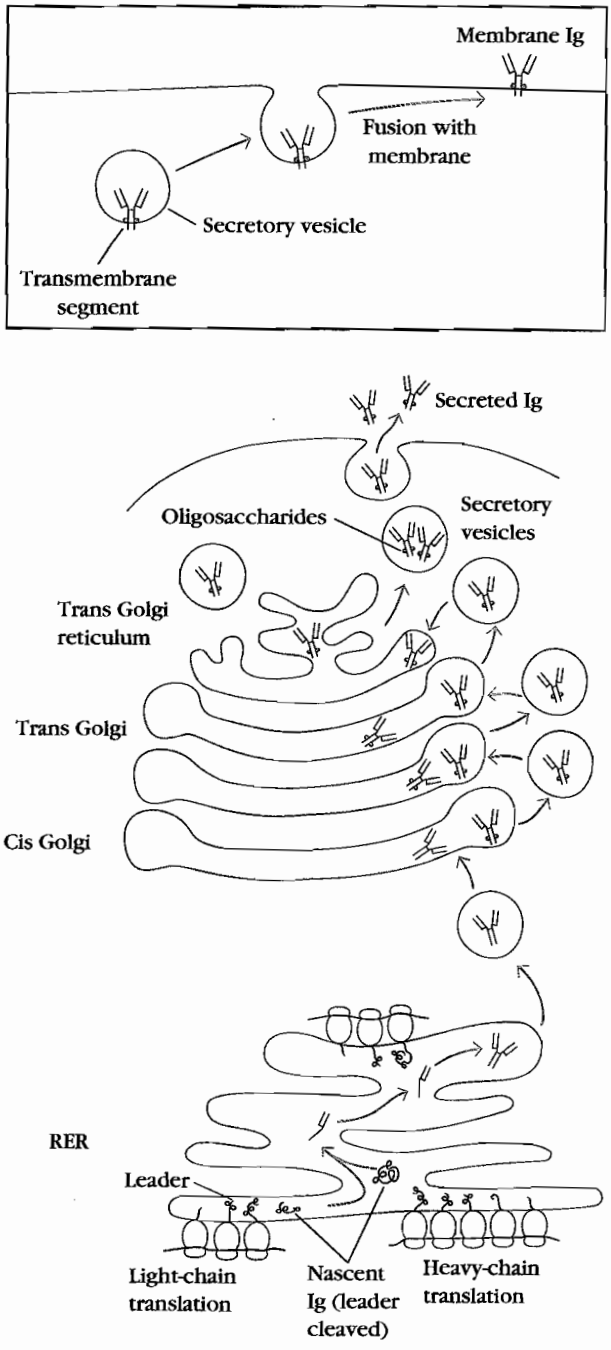


FIGURE 5-18 Synthesis, assembly, and secretion of the immunoglobulin molecule. The heavy and light chains are synthesized on separate polyribosomes (polysomes). The assembly of the chains to form the disulfide-linked immunoglobulin molecule occurs as the chains pass through the cisternae of the rough endoplasmic reticulum (RER) into the Golgi apparatus and then into secretory vesicles. The main figure depicts assembly of a secreted antibody. The inset depicts a membrane-bound antibody, which contains the carboxyl-terminal transmembrane segment. This form becomes anchored in the membrane of secretory vesicles and then is inserted into the cell membrane when the vesicles fuse with the membrane.

Regulation of Ig-Gene Transcription

The immunoglobulin genes are expressed only in B-lineage cells, and even within this lineage, the genes are expressed at different rates during different developmental stages. As with other eukaryotic genes, three major classes of cis regulatory sequences in DNA regulate transcription of immunoglobulin genes:

- **Promoters:** relatively short nucleotide sequences, extending about 200 bp upstream from the transcription initiation site, that promote initiation of RNA transcription in a specific direction
- **Enhancers:** nucleotide sequences situated some distance upstream or downstream from a gene that activate transcription from the promoter sequence in an orientation-independent manner
- **Silencers:** nucleotide sequences that down-regulate transcription, operating in both directions over a distance.

The locations of the three types of regulatory elements in germ-line immunoglobulin DNA are shown in Figure 5-19. All of these regulatory elements have clusters of sequence motifs that can bind specifically to one or more nuclear proteins.

Each V_H and V_L gene segment has a promoter located just upstream from the leader sequence. In addition, the Jκ cluster and each of the D_H genes of the heavy-chain locus are preceded by promoters. Like other promoters, the immunoglobulin promoters contain a highly conserved AT-rich sequence called the TATA box, which serves as a site for the binding of a number of proteins that are necessary for the initiation of RNA transcription. The actual process of transcription is performed by RNA polymerase II, which starts transcribing DNA from the initiation site, located about 25 bp downstream of the TATA box. Ig promoters also contain an essential and conserved octamer that confers B-cell specificity on the promoter. The octamer binds two transcription factors, oct-1, found in many cell types, and oct-2, found only in B cells.

While much remains to be learned about the function of enhancers, they have binding sites for a number of proteins, many of which are transcription factors. A particularly important role is played by two proteins encoded by the *E2A* gene which can undergo alternate splicing to generate two collaborating proteins, both of which bind to the μ and κ intronic enhancers. These proteins are essential for B-cell development and E2A knockout mice make normal numbers of T cells but show a total absence of B cells. Interestingly, transfection of these enhancer-binding proteins into a T cell line resulted in a dramatic increase in the transcription of μ chain mRNA and even induced the T cell to undergo $D_H + J_H \rightarrow D_HJ_H$ rearrangement. Silencers may inhibit the activity of Ig

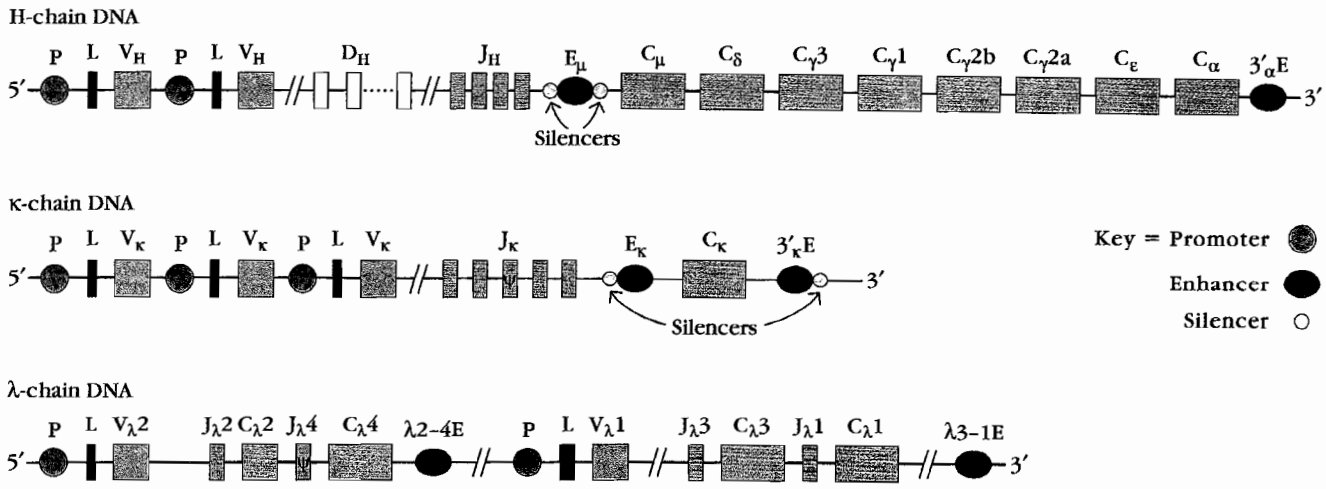


FIGURE 5-19 Location of promoters (dark red), enhancers (green), and silencers (yellow) in mouse heavy-chain, κ light-chain, and λ light-chain germ-line DNA. Variable-region DNA rearrangement moves an enhancer close enough to a promoter that the enhancer can activate transcription from the promoter. The promoters that precede the DH cluster, a number of the C genes and the J_λ cluster are omitted from this diagram for the sake of clarity.

enhancers in non-B cells. If so, they could be important contributors to the high levels of Ig gene transcription that are characteristic of B cells but absent in other cell types.

One heavy-chain enhancer is located within the intron between the last (3′) J gene segment and the first (5′) C gene segment (C_μ), which encodes the μ heavy chain. Because this heavy-chain enhancer (E_μ) is located 5′ of the S_μ switch site near C_μ, it can continue to function after class switching has occurred. Another heavy-chain enhancer ($3'_\alpha E$) has been detected 3′ of the C_α gene segment. One κ light-chain enhancer (E_κ) is located between the J_κ segment and the C_κ segment, and another enhancer ($3'_\kappa E$) is located 3′ of the C_κ segment. The λ light-chain enhancers are located 3′ of $C_\lambda4$ and 3′ of $C_\lambda1$. Silencers have been identified in heavy-chain and κ-chain DNA, adjacent to enhancers, but not in λ-chain DNA.

DNA Rearrangement Greatly Accelerates Transcription

The promoters associated with the immunoglobulin V gene segments bind RNA polymerase II very weakly, and the variable-region enhancers in germ-line DNA are quite distant from the promoters (about 250–300 kb), too remote to significantly influence transcription. For this reason, the rate of transcription of V_H and V_L coding regions is negligible in unrearranged germ-line DNA. Variable-region gene rearrangement brings a promoter and enhancer within 2 kb of each other, close enough for the enhancer to influence transcription from the nearby promoter. As a result, the rate of transcription of a rearranged $V_L J_L$ or $V_H D_H J_H$ unit is as much as 10^4 times the rate of transcription of unrearranged V_L or V_H segments. This effect was demonstrated directly in a study in which B cells transfected with rearranged heavy-chain genes from which the enhancer had been deleted did not transcribe the genes, whereas B cells transfected with similar genes that contained the enhancer transcribed the transfected genes at a high rate. These findings highlight the importance of enhancers in the normal transcription of immunoglobulin genes.

Genes that regulate cellular proliferation or prohibit cell death sometimes translocate to the immunoglobulin heavy- or light-chain loci. Here, under the influence of an immunoglobulin enhancer, the expression of these genes is significantly elevated, resulting in high levels of growth promoting or cell death inhibiting proteins. Translocations of the c-*myc* and *bcl*-2 oncogenes have each been associated with malignant B-cell lymphomas. The translocation of c-*myc* leads to constitutive expression of c-Myc and an aggressive, highly proliferative B-cell lymphoma called Burkitt's lymphoma. The translocation of *bcl*-2 leads to suspension of programmed cell death in B cells, resulting in follicular B-cell lymphoma. These cancer-promoting translocations are covered in greater detail in Chapter 22.

Ig-Gene Expression Is Inhibited in T Cells

As noted earlier, germ-line DNA encoding the T-cell receptor (TCR) undergoes V-(D)-J rearrangement to generate functional TCR genes. Rearrangement of both immunoglobulin and TCR germ-line DNA occurs by similar recombination processes mediated by RAG-1 and RAG-2 and involving recombination signal sequences with one-turn or two-turn spacers (see Figure 5-7). Despite the similarity of the processes, complete Ig-gene rearrangement of H and L chains occurs only in B cells and complete TCR-gene rearrangement is limited to T cells.

Hitoshi Sakano and coworkers have obtained results suggesting that a sequence within the κ-chain 3′ enhancer ($3'_{\kappa}E$) serves to regulate the joining of V_{κ} to J_{κ} in B and T cells. When a sequence known as the PU.1 binding site within the 3′ κ-chain enhancer was mutated, these researchers found that V_{κ}-J_{κ} joining occurred in T cells as well as B cells. They propose that binding of a protein expressed by T cells, but not B cells, to the unmutated κ-chain enhancer normally prevents V_{κ}-J_{κ} joining in T cells. The identity of this DNA-binding protein in T cells remains to be determined. Similar processes may prevent rearrangement of heavy-chain and λ-chain DNA in T cells.

Antibody Genes and Antibody Engineering

There are many clinical applications in which the exquisite specificity of a mouse monoclonal antibody would be useful. However, when mouse monoclonal antibodies are introduced into humans they are recognized as foreign and evoke an antibody response that quickly clears the mouse monoclonal antibody from the bloodstream. In addition, circulating complexes of mouse and human antibodies can cause allergic reactions. In some cases, the buildup of these complexes in organs such as the kidney can cause serious and even life-threatening reactions. Clearly, one way to avoid these undesirable reactions is to use human monoclonal antibodies for clinical applications. However, the preparation of human monoclonal antibodies has been hampered by numerous technical problems. In response to the difficulty of producing human monoclonal antibodies and the complications resulting from the use of mouse monoclonal antibodies in humans, there is now a major effort to engineer monoclonal antibodies and antibody binding sites with recombinant DNA technology.

The growing knowledge of antibody gene structure and regulation has made possible what Cesar Milstein, one of the inventors of monoclonal antibody technology, has called "man-made antibodies." It is now possible to design and construct genes that encode immunoglobulin molecules in which the variable regions come from one species and the constant regions come from another. New genes have been created that link nucleotide sequences coding nonantibody proteins with sequences that encode antibody variable regions specific for particular antigens. These molecular hybrids or **chimeras** may be able to deliver powerful toxins to particular antigenic targets, such as tumor cells. Finally, by replacement of the immunoglobulin loci of one species with that of another, animals of one species have been endowed with the capacity to respond to immunization by producing antibodies encoded by the donor's genetically transplanted Ig genes. By capturing a significant sample of all of the immunoglobulin heavy- and light-chain variable-region genes via incorporation into libraries of bacteriophage, it has been possible to achieve significant and useful reconstructions of the entire antibody repertoires of individuals. The next few sections describe each of these types of antibody genetic engineering.

Chimeric and Hybrid Monoclonal Antibodies Have Potent Clinical Potential

One approach to engineering an antibody is to clone recombinant DNA containing the promoter, leader, and variable-region sequences from a mouse antibody gene and the constant-region exons from a human antibody gene (Figure 5-20). The antibody encoded by such a recombinant gene is a mouse-human chimera, commonly known as a **humanized antibody.** Its antigenic specificity, which is determined by the variable region, is derived from the mouse DNA; its isotype, which is determined by the constant region, is derived from the human DNA. Because the constant regions of these chimeric antibodies are encoded by human genes, the anti-

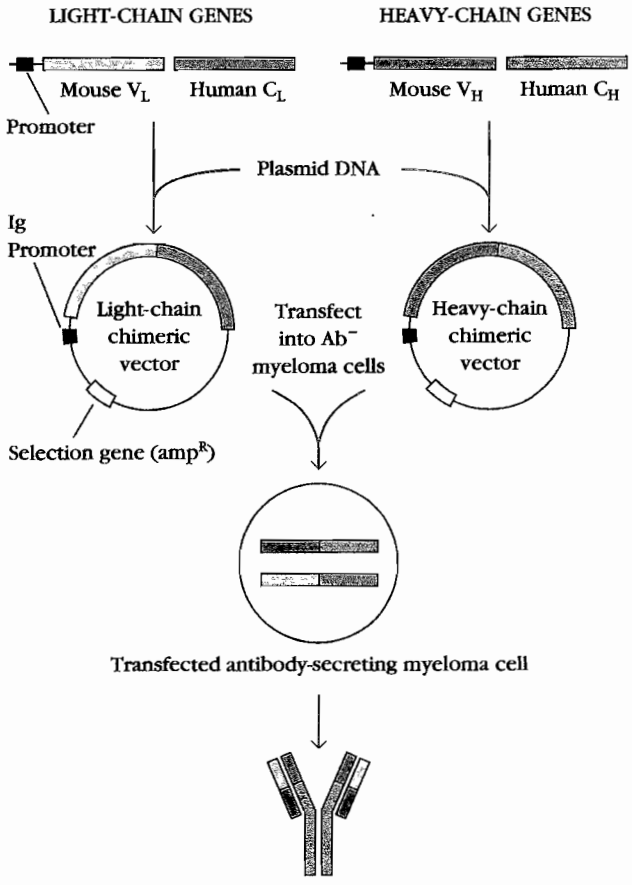


FIGURE 5-20 Production of chimeric mouse-human monoclonal antibodies. Chimeric mouse-human heavy- and light-chain expression vectors are produced. These vectors are transfected into Ab^- myeloma cells. Culture in ampicillin medium selects for transfected myeloma cells that secrete the chimeric antibody. *[Adapted from M. Verhoeyen and L. Reichmann, 1988,* BioEssays ***8:****74.]*

bodies have fewer mouse antigenic determinants and are far less immunogenic when administered to humans than mouse monoclonal antibodies (Figure 5-21a). The ability of the mouse variable regions remaining in these humanized antibodies to provide the appropriate binding site to allow specific recognition of the target antigen has encouraged further exploration of this approach. It is possible to produce chimeric human-mouse antibodies in which only the sequences of the CDRs are of mouse origin (Figure 5-21b). Another advantage of humanized chimeric antibodies is that they retain the biological effector functions of human antibody and are more likely to trigger human complement activation or Fc receptor binding. One such chimeric human-mouse antibody has been used to treat patients with B-cell varieties of non-Hodgkin's lymphoma (see Clinical Focus).

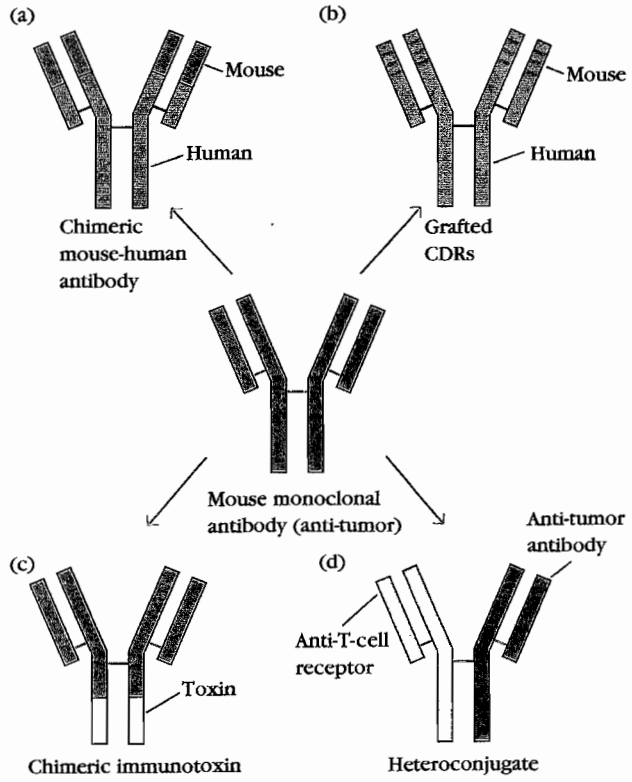


FIGURE 5-21 Chimeric and hybrid monoclonal antibodies engineered by recombinant DNA technology. (a) Chimeric mouse-human monoclonal antibody containing the V_H and V_L domains of a mouse monoclonal antibody (blue) and C_L and C_H domains of a human monoclonal antibody (gray). (b) A chimeric monoclonal antibody containing only the CDRs of a mouse monoclonal antibody (blue bands) grafted within the framework regions of a human monoclonal antibody is called a "humanized" monoclonal antibody. (c) A chimeric monoclonal antibody in which the terminal Fc domain is replaced by toxin chains (white). (d) A heteroconjugate in which one-half of the mouse antibody molecule is specific for a tumor antigen and the other half is specific for the CD3/T-cell receptor complex.

Chimeric monoclonal antibodies that function as immunotoxins (see Figure 4-23) can also be prepared. In this case, the terminal constant-region domain in a tumor-specific monoclonal antibody is replaced with toxin chains (Figure 5-21c). Because these immunotoxins lack the terminal Fc domain, they are not able to bind to cells bearing Fc receptors. These immunotoxins can bind only to tumor cells, making them highly specific as therapeutic reagents.

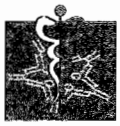
Heteroconjugates, or **bispecific antibodies,** are hybrids of two different antibody molecules (Figure 5-21d). They can be constructed by chemically crosslinking two different antibodies or by synthesizing them in hybridomas consisting of two different monoclonal-antibody-producing cell lines that have been fused. Both of these methods generate mixtures of monospecific and bispecific antibodies from which the desired bispecific molecule must be purified. Using genetic engineering to construct genes that will encode molecules only with the two desired specificities is a much simpler and more elegant approach. Several bispecific molecules have been designed in which one half of the antibody has specificity for a tumor and the other half has specificity for a surface molecule on an immune effector cell, such as an NK cell, an activated macrophage, or a cytotoxic T lymphocyte (CTL). Such heteroconjugates have been designed to activate the immune effector cell when it is crosslinked to the tumor cell so that it begins to mediate destruction of the tumor cell.

Monoclonal Antibodies Can Be Constructed from Ig-Gene Libraries

A quite different approach for generating monoclonal antibodies employs the polymerase chain reaction (PCR) to amplify the DNA that encodes antibody heavy-chain and light-chain Fab fragments from hybridoma cells or plasma cells. A promoter region and *Eco*RI restriction site (see Chapter 23) are added to the amplified sequences, and the resulting constructs are inserted into bacteriophage λ, yielding separate heavy- and light-chain libraries. Cleavage with *Eco*RI and random joining of the heavy- and light-chain genes yield numerous novel heavy-light constructs (Figure 5-22).

This procedure generates an enormous diversity of antibody specificities—libraries with $>10^{10}$ unique members have been obtained—and clones containing these random combinations of H + L chains can be rapidly screened for those secreting antibody to a particular antigen. The level of diversity is comparable to the human in vivo repertoire, and it is possible to demonstrate that specificities against a wide variety of antigens can be obtained from these libraries. Such a combinatorial library approach opens the possibility of obtaining specific antibodies without any need whatsoever for immunization.

However, the real challenge to bypassing in vivo immunization in the derivation of useful antibodies of high affinity lies in finding ways to mimic the biology of the humoral



CLINICAL FOCUS

Therapy for Non-Hodgkin's Lymphoma and Other Diseases by Genetically Engineered Antibodies

Lymphomas are cancers of lymphatic tissue in which the tumor cells are of lymphocytic origin. There are two major forms of lymphoma: Hodgkin's lymphoma and non-Hodgkin's lymphoma. The less common form is Hodgkin's lymphoma, named for its discoverer, Thomas Hodgkin, an English physician. This unusually gifted early pathologist, who worked without the benefit of a microscope, recognized this condition in several patients and first described the anatomical features of the disease in 1832. Because many tissue specimens taken from patients Hodgkin suspected of harboring the disease were saved in the Gordon Museum of Guy's Hospital in London, it has been possible for later generations to judge the accuracy of his diagnoses. Hodgkin has fared well. Studies of these preserved tissues confirm that he was right in about 60% of the cases, a surprising achievement, considering the technology of the time. Actually, most lymphoma is non-Hodgkin's type and includes about 10 different types of disease. B-cell lymphomas are an important fraction of these.

For some years now, the major therapies directed against lymphomas have been radiation, chemotherapy, or a combination of both. While these therapies benefit large numbers of patients by increasing survival, relapses after treatment are common, and many treated patients experience debilitating side effects. The side effects are an expected consequence of these therapies, because the agents used kill or severely damage a broad spectrum of normal cells as well as tumor cells. One of the holy grails of cancer treatment is the discovery of therapies that will affect only the tumor cells and completely spare normal cells. If particular types of cancer cells had antigens that were tumor specific, these antigens would be ideal targets for immune attack. Unfortunately, there are few such molecules known. However, a number of antigens are known that are restricted to the cell lineage in which the tumor originated and are expressed on the tumor cells.

Many cell-lineage-specific antigens have been identified for B lymphocytes and B lymphomas, including immunoglobulin, the hallmark of the B cell, and CD20, a membrane-bound phosphoprotein. CD20 has emerged as an attractive candidate for antibody-mediated immunotherapy because it is present on B lymphomas, and antibody-mediated crosslinking does not cause it to downregulate or internalize. Indeed, some years ago, mouse monoclonal antibodies were raised against CD20, and one of these has formed the basis for an anti-B-cell lymphoma immunotherapy. This approach appears ready to take its place as an adjunct or alternative to radiation and chemotherapy. The development of this anti-tumor antibody is an excellent case study of the combined application of immunological insights and molecular biology to engineer a novel therapeutic agent.

The original anti-CD20 antibody was a mouse monoclonal antibody with murine γ heavy chains and κ light chains. The DNA sequences of the light- and heavy-chain variable regions of this antibody were amplified by PCR. Then a chimeric gene was created by replacing the CDR gene sequences of a human $\gamma 1$ heavy chain with those from the murine heavy chain. In a similar maneuver, CDRs from the mouse κ were ligated into a human κ gene. The chimeric genes thus created were incorporated into vectors that permitted high levels of expression in mammalian cells. When an appropriate cell line was co-transfected with both of these constructs, it produced chimeric antibodies containing CDRs of mouse origin together with human variable-region frameworks and constant regions. After purification, the biological activity of the antibody was evaluated, first in vitro and then in a primate animal model.

The initial results were quite promising. The grafted human constant region supported effector functions such as the complement-mediated lysis or antibody-dependent cell-mediated cytotoxicity (ADCC) of human B lymphoid cells. Furthermore, weekly injections of the antibody into monkeys resulted in the rapid and sustained depletion of B cells from peripheral blood, lymph nodes, and even bone marrow. When the anti-CD20 antibody infusions were stopped, the differentiation of new B cells from progenitor populations allowed B-cell populations eventually to recover and approach normal levels. From these results, the hope grew that this immunologically active chimeric antibody could be used to clear entire B cell populations, including B lymphoma cells, from the body in a way that spared other cell populations. This led to the trial of the antibody in human patients.

The human trials enrolled patients with B-cell lymphoma who had a relapse after chemotherapy or radiation treatment. These trials addressed three important issues: efficacy, safety, and immunogenicity. While not all patients responded to treatment with anti-CD20, close to 50% exhibited full or partial remission. Thus, efficacy was demonstrated, because this level of response is comparable to the success rate with traditional approaches that employ highly cytotoxic drugs or radiation—it offers a truly alternative therapy. Side effects such as nausea, low blood pressure, and shortness of breath were seen in some patients (usually during or shortly after the initiation of therapy); these were, for the most part, not serious or life-threatening. Consequently, treatment with the

chimeric anti-CD20 appears safe. Patients who received the antibody have been observed closely for the appearance of human anti-mouse-Ig antibodies (HAMA) and for human anti-chimeric antibody (HACA) responses. Such responses were not observed. Therefore, the antibody was not immunogenic. The absence of such responses demonstrate that antibodies can be genetically engineered to minimize, or even avoid, untoward immune reactions. Another reason for humanizing mouse antibodies arises from the very short half life (a few hours) of mouse IgG antibodies in humans compared with the three-week half lives of their human or humanized counterparts.

Antibody engineering has also contributed to the therapy of other malignancies such as breast cancer, which is diagnosed in more than 180,000 American women each year. A little more than a quarter of all breast cancer patients have cancers that over-express a growth factor receptor called HER2 (human epidermal growth factor receptor 2). Many tumors that over-express HER2 grow faster and pose a more serious threat than those with normal levels of this protein on their surface. A chimeric anti-HER2 monoclonal antibody in which all of the protein except the CDRs are of human origin was created by genetic engineering. Specifically, the DNA sequences for the heavy-chain and light-chain CDRs were taken from cloned mouse genes encoding an anti-HER2 monoclonal antibody. As in the anti-CD20 strategy described above, each of the mouse CDR gene segments were used to replace the corresponding human CDR gene segments in human genes encoding the human IgG_1 heavy chain and the human κ light chain. When this engineered antibody is used in combination with a chemotherapeutic drug, it is highly effective against metastatic breast cancer. The effects on patients who were given only a chemotherapeutic drug were compared with those for patients receiving both the chemotherapeutic drug and the engineered anti-HER2 antibody. The combination anti-HER2/chemotherapy treatment showed significantly reduced rates of tumor progression, a higher percentage of responding patients, and a higher one-year survival rate. Treatment with Herceptin, as this engineered monoclonal antibody is called, has become part of the standard repertoire of breast cancer therapies.

The development of engineered and conventional monoclonal antibodies is one of the most active areas in the pharmaceutical industry. The table provides a partial compilation of monoclonal antibodies that have received approval from the Food and Drug Administration (FDA) for use in the treatment of human disease. Many more are in various stages of development and testing.

Some monoclonal antibodies in clinical use

Monoclonal antibody [mAB] (Product Name)	Nature of antibody	Target (antibody specificity)	Treatment for
Muromonab-CD3 (Orthoclone OKT3)	Mouse mAB	T cells (CD3, a T cell antigen)	Acute rejection of liver, heart and kidney transplants
Abciximab (ReoPro)	Human-mouse chimeric	Clotting receptor of platelets (GP IIb/IIIa)	Blood clotting during angioplasty and other cardiac procedures
Daclizumab (Zenapax)	Humanized mAB	Activated T cells (IL-2 receptor alpha subunit)	Acute rejection of kidney transplants
Inflixibmab (Remicade)	Human-mouse chimeric	Tumor necrosis factor, (TNF) a mediator of inflammation. (TNF)	Rheumatoid arthritis and Crohn's disease
Palivizumab (Synagis)	Humanized mAB	Respiratory Syncytial Virus (RSV) (F protein, a component of RSV)	RSV infection in children, particularly infants
Gemtuzumab (Mylotarg)	Humanized mAB	Many cells of the myeloid lineage (CD33, an adhesion molecule)	Acute myeloid leukemia (AML)
Alemtuzumab (Campath)	Humanized mAB	Many types of leukocytes (CD52 a cell surface antigen)	B cell chronic lymphocytic leukemia
Trastuzumab (Herceptin)	Humanized mAB	An epidermal growth factor receptor (HER2 receptor)	HER2 receptor-positive advanced breast cancers
Rituximab (Rituxan)	Humanized mAB	B cells (CD20 a B cell surface antigen)	Relapsed or refractory non-Hodgkins lymphoma
Ibritumomab (Zevalin)	Mouse mAB	B cells (CD20, a B cell surface antigen)	Relapsed or refractory non-Hodgkins lymphoma

SOURCE: Adapted from P. Carter. 2001. Improving the efficacy of antibody-based cancer therapies. *Nature Reviews/Cancer* 1:118.

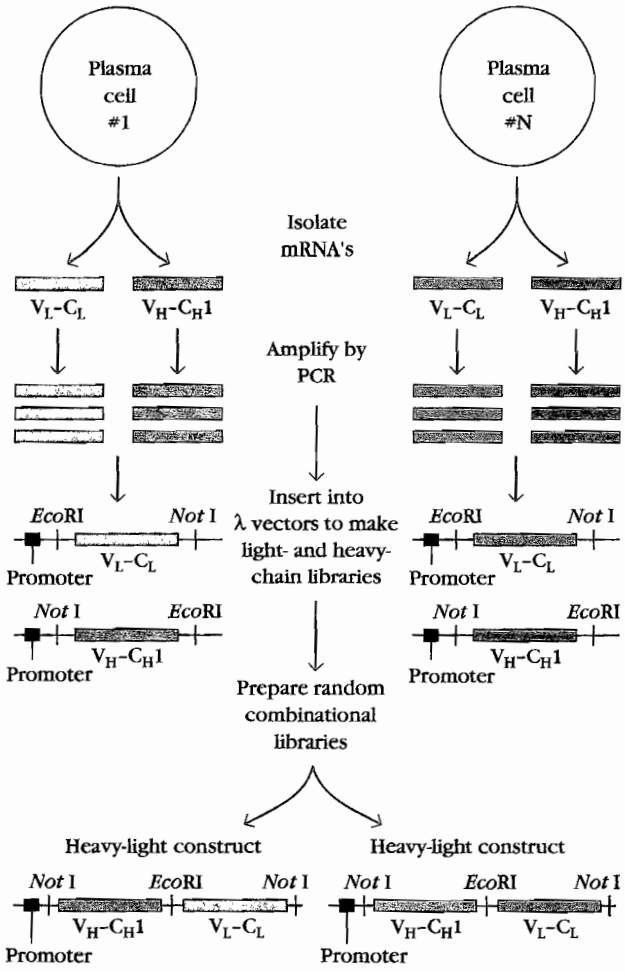


FIGURE 5-22 General procedure for producing gene libraries encoding Fab fragments. In this procedure, isolated mRNA that encodes heavy and light chains is amplified by the polymerase chain reaction (PCR) and cloned in λ vectors. Random combinations of heavy- and light-chain genes generate an enormous number of heavy-light constructs encoding Fab fragments with different antigenic specificity. *[Adapted from W. D. Huse et al., 1989,* Science ***246:****1275.]*

immune response. As we shall see in Chapter 11, the in vivo evolution of most humoral immune responses produces two desirable outcomes. One is class switching, in which a variety of antibody classes of the same specificity are produced. This is an important consideration because the class switching that occurs during an immune response produces antibodies that have the same specificity but different effector functions and hence, greater biological versatility. The other is the generation of antibodies of higher and higher affinity as the response progresses. A central goal of Ig-gene library approaches is the development of strategies to produce antibodies of appropriate affinity in vitro as readily as they are generated by an in vivo immune response. When the formidable technical obstacles to the achievement of these goals are overcome, combinatorial approaches based on phage libraries will allow the routine and widespread production of useful antibodies from any desired species without the limitations of immunization and hybridoma technology that currently complicate the production of monoclonal antibodies.

Mice Have Been Engineered with Human Immunoglobulin Loci

It is possible to functionally knock out, or disable, the heavy- and light-chain immunoglobulin loci in mouse embryonic stem (ES) cells. N. Lonberg and his colleagues followed this procedure and then introduced large DNA sequences (as much as 80 kb) containing human heavy- and light-chain gene segments. The DNA sequences contained constant-region gene segments, J segments, many V-region segments, and, in the case of the heavy chain, D_H segments. The ES cells containing these miniature human Ig gene loci (miniloci) are used to derive lines of transgenic mice that respond to antigenic challenge by producing antigen-specific *human* antibodies (Figure 5-23). Because the human heavy- and light-chain miniloci undergo rearrangement and all the other diversity-generating processes, such as N-addition, P-addition, and even somatic hypermutation after antigenic challenge, there is an opportunity for the generation of a great deal of diversity in these mice. The presence of human heavy-chain minilocus genes for more than one isotype and their accompanying switch sites allows class switching as well. A strength of this method is that these completely human antibodies are made in cells of the mouse B-cell lineage, from which antibody-secreting hybridomas are readily derived by cell fusion. This approach thus offers a solution to the problem of producing human monoclonal antibodies of any specificity desired.

SUMMARY

- Immunoglobulin κ and λ light chains and heavy chains are encoded by three separate multigene families, each containing numerous gene segments and located on different chromosomes.
- Functional light-chain and heavy-chain genes are generated by random rearrangement of the variable-region gene segments in germ-line DNA.
- V(D)J joining is catalyzed by the recombinase activiating genes, *RAG-1* and *RAG-2,* and the participation of other enzymes and proteins. The joining of segments is directed by recombination signal sequences (RSS), conserved DNA sequences that flank each V, D, and J gene segment.
- Each recombination signal sequence contains a conserved heptamer sequence, a conserved nonamer sequence, and either a 12-bp (one-turn) or 23-bp (two-turn) spacer. During rearrangement, gene segments flanked by a one-turn spacer join only to segments flanked by a two-turn spacer, assuring proper V_L-J_L and V_H-D_H-J_H joining.

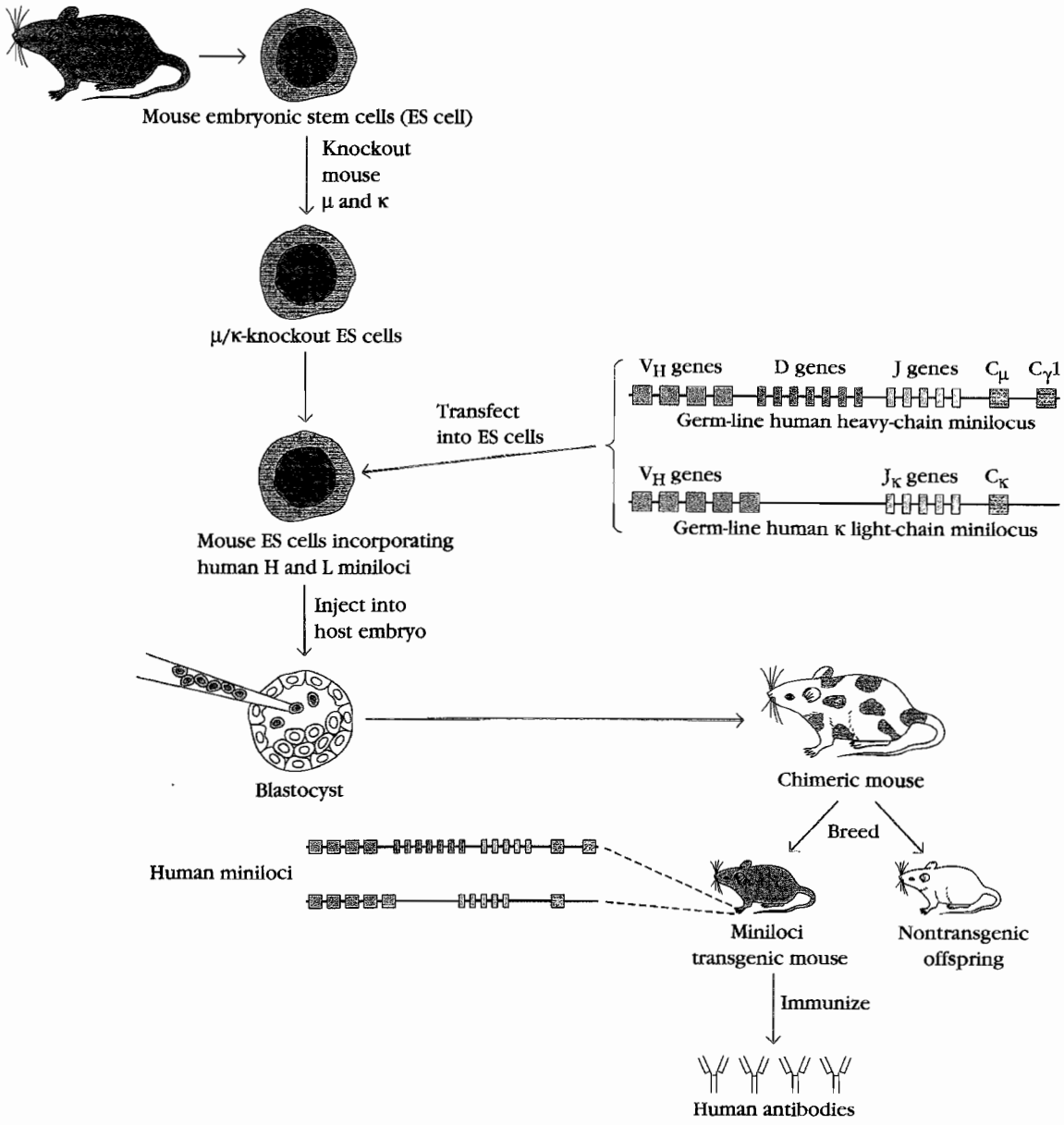


FIGURE 5-23 Grafting human heavy- and light-chain miniloci into mice. The capacity of mice to rearrange Ig heavy- and light-chain gene segments was disabled by knocking out the C_μ and C_κ loci. The antibody-producing capacity of these mice was reconstituted by introducing long stretches of DNA incorporating a large part of the human germ-line κ and heavy-chain loci (miniloci). Chimeric mice were then bred to establish a line of transgenic mice bearing both heavy- and light-chain human miniloci. Immunization of these mice results in the production of human antibody specific for the target antigen. *[N. Lonberg et al., 1994,* Nature ***368:****856.]*

- Immunoglobulin gene rearrangements occur in sequential order, heavy-chain rearrangements first, followed by light-chain rearrangements. Allelic exclusion is a consequence of the functional rearrangement of the immunoglobulin DNA of only one parental chromosome and is necessary to assure that a mature B cell expresses immunoglobulin with a single antigenic specificity.
- The major sources of antibody diversity, which can generate $>10^{10}$ possible antibody combining sites, are: random joining of multiple V, J, and D germ-line gene segments; random association of heavy and light chains; junctional flexibility; P-addition; N-addition; and somatic mutation.
- After antigenic stimulation of mature B cells, class switching results in expression of different classes of antibody (IgG, IgA, and IgE) with the same antigenic specificity.
- Differential RNA processing of the immunoglobulin heavy-chain primary transcript generates membrane-bound antibody in mature B cells, secreted antibody in

plasma cells, and the simultaneous expression of IgM and IgD by mature B cells.

- Transcription of immunoglobulin genes is regulated by three types of DNA regulatory sequences: promoters, enhancers, and silencers.
- Growing knowledge of the molecular biology of immunoglobulin genes has made it possible to engineer antibodies for research and therapy. The approaches include chimeric antibodies, bacteriophage-based combinatorial libraries of Ig-genes, and the transplantation of extensive segments of human Ig loci into mice.

References

Chen, J., Y. Shinkai, F. Young, and F. W. Alt. 1994. Probing immune functions in RAG-deficient mice. *Curr. Opin. Immunol.* **6:**313.

Cook, G. P., and I. M. Tomlinson. 1995. The human immunoglobulin V_H repertoire. *Immunol. Today* **16:**237.

Dreyer, W. J., and J. C. Bennett. 1965. The molecular basis of antibody formation. *Proc. Natl. Acad. Sci. U.S.A.* **54:**864.

Fugmann, S. D., I. L. Lee, P. E. Shockett, I. J. Villey, and D. G. Schatz. 2000. The RAG proteins and V(D)J recombination: Complexes, ends and transposition. *Annu. Rev. Immunol.* **18:**495.

Gavilondo, J. V., and J. W. Larrick. 2000. Antibody engineering at the millennium. *Biotechniques* **29:**128.

Hayden, M. S., L. K. Gilliand, and J. A. Ledbetter. 1997. Antibody engineering. *Curr. Opin. Immunol.* **9:**201.

Hesslein, D. G., and D. G. Schatz. 2001. Factors and forces controlling V(D)J recombination. *Adv. Immunol.* **78:**169.

Hozumi, N., and S. Tonegawa. 1976. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. U.S.A.* **73:**3628.

Maloney, D. G., et al. 1997. IDEC-C2B8 (Rituximab) anti-CD20 monoclonal antibody therapy in patients with relapsed low-grade non-Hodgkin's lymphoma. *Blood* **90:**2188.

Manis, J. P., M. Tian, and F. W. Alt. 2002. Mechanism and control of class-switch recombination. *Trends Immunol.* **23:**31.

Matsuda, F., K. Ishii, P. Bourvagnet, Ki Kuma, H. Hayashida, T. Miyata, and T. Honjo. 1998. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J. Exp. Med.* **188:**2151.

Max, E. E. 1998. Immunoglobulins: molecular genetics. In *Fundamental Immunology,* 4th ed., W. E. Paul, ed. Lippincott-Raven, Philadelphia.

Mills, F. C., N. Harindranath, M. Mitchell, and E. E. Max. 1997. Enhancer complexes located downstream of both human immunoglobulin C alpha genes. *J. Exp. Med.* **186:**845.

Oettinger, M. A., et al. 1990. *RAG-1* and *RAG-2,* adjacent genes that synergistically activate V(D)J recombination. *Science* **248:**1517.

Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature* **302:**575.

Van Gent, D. C., et al. 1995. Initiation of V(D)J recombination in a cell-free system. *Cell* **81:**925.

Winter, G., and C. Milstein. 1990. Man-made antibodies. *Nature* **349:**293.

USEFUL WEB SITES

http://www.mrc-cpe.cam.ac.uk/imt-doc/public/INTRO.html#maps

V BASE: This database and informational site is maintained at the MRC Centre for Protein Engineering in England. It is an excellent and comprehensive directory of information on the human germ-line variable region.

http://www.mgen.uni-heidelberg.de/SD/SDscFvSite.html

The Recombinant Antibody Page: This site has a number of links that provide interesting opportunities to explore the potential of genetic engineering of antibodies.

http://www.ebi.ac.uk/imgt/hla/intro.html.

The IMGT site contains a collection of databases of genes relevant to the immune system. The IMGT/LIGM database houses sequences belonging to the immunoglobulin superfamily and of T cell antigen receptor sequences.

Study Questions

CLINICAL FOCUS QUESTION The Clinical Focus section includes a table of monoclonal antibodies approved for clinical use. Two, Rituxan and Zevalin, are used for the treatment of non-Hodgkins lymphoma. Both target CD20, a B-cell surface antigen. Zevalin is chemically modified by attachment of radioactive isotopes (yttrium-90, a β emitter or indium-111, a high energy γ emitter) that lethally irradiate cells to which the monoclonal antibody binds. Early experiments found that Zevalin without a radioactive isotope attached was an ineffective therapeutic agent, whereas unlabeled Rituxan, a humanized mAB, was effective. Furthermore, Rituxan with a radioactive isotope attached was too toxic; Zevalin bearing the same isotope in equivalent amounts was far less toxic. Explain these results. (Hint: The longer a radioactive isotope stays in the body, the greater the dose of radiation absorbed by the body.)

1. Indicate whether each of the following statements is true or false. If you think a statement is false, explain why.
   a. V_κ gene segments sometimes join to C_λ gene segments.
   b. With the exception of a switch to IgD, immunoglobulin class switching is mediated by DNA rearrangements.
   c. Separate exons encode the transmembrane portion of each membrane immunoglobulin.
   d. Although each B cell carries two alleles encoding the immunoglobulin heavy and light chains, only one allele is expressed.

e. Primary transcripts are processed into functional mRNA by removal of introns, capping, and addition of a poly-A tail.
f. The primary transcript is an RNA complement of the coding strand of the DNA and includes both introns and exons.

2. Explain why a V_H segment cannot join directly with a J_H segment in heavy-chain gene rearrangement.

3. Considering only combinatorial joining of gene segments and association of light and heavy chains, how many different antibody molecules potentially could be generated from germ-line DNA containing 500 V_L and 4 J_L gene segments and 300 V_H, 15 D_H, and 4 J_H gene segments?

4. For each incomplete statement below (a–g), select the phrase(s) that correctly completes the statement. More than one choice may be correct.

a. Recombination of immunoglobulin gene segments serves to
(1) promote Ig diversification
(2) assemble a complete Ig coding sequence
(3) allow changes in coding information during B-cell maturation
(4) increase the affinity of immunoglobulin for antibody
(5) all of the above

b. Somatic mutation of immunoglobulin genes accounts for
(1) allelic exclusion
(2) class switching from IgM to IgG
(3) affinity maturation
(4) all of the above
(5) none of the above

c. The frequency of somatic mutation in Ig genes is greatest during
(1) differentiation of pre-B cells into mature B cells
(2) differentiation of pre-T cells into mature T cells
(3) generation of memory B cells
(4) antibody secretion by plasma cells
(5) none of the above

d. Kappa and lambda light-chain genes
(1) are located on the same chromosome
(2) associate with only one type of heavy chain
(3) can be expressed by the same B cell
(4) all of the above
(5) none of the above

e. Generation of combinatorial diversity among immunoglobulins involves
(1) mRNA splicing
(2) DNA rearrangement
(3) recombination signal sequences
(4) one-turn/two-turn joining rule
(5) switch sites

f. A B cell becomes immunocompetent
(1) following productive rearrangement of variable-region heavy-chain gene segments in germ-line DNA
(2) following productive rearrangement of variable-region heavy-chain and light-chain gene segments in germ-line DNA
(3) following class switching
(4) during affinity maturation
(5) following binding of T_H cytokines to their receptors on the B cell

g. The mechanism that permits immunoglobulins to be synthesized in either a membrane-bound or secreted form is
(1) allelic exclusion
(2) codominant expression
(3) class switching
(4) the one-turn/two-turn joining rule
(5) differential RNA processing

5. What mechanisms generate the three hypervariable regions (complementarity-determining regions) of immunoglobulin heavy and light chains? Why is the third hypervariable region (CDR3) more variable than the other two (CDR1 and CDR2)?

6. You have been given a cloned myeloma cell line that secretes IgG with the molecular formula $\gamma_2\lambda_2$. Both the heavy and light chains in this cell line are encoded by genes derived from allele 1. Indicate the form(s) in which each of the genes listed below would occur in this cell line using the following symbols: G = germ line form; R = productively rearranged form; NP = nonproductively rearranged form. State the reason for your choice in each case.

a. Heavy-chain allele 1
b. Heavy-chain allele 2
c. κ-chain allele 1
d. κ-chain allele 2
e. λ-chain allele 1
f. λ-chain allele 2

7. You have a B-cell lymphoma that has made nonproductive rearrangements for both heavy-chain alleles. What is the arrangement of its light-chain DNA? Why?

8. Indicate whether each of the class switches indicated below can occur (Yes) or cannot occur (No).

a. IgM to IgD
b. IgM to IgA
c. IgE to IgG
d. IgA to IgG
e. IgM to IgG

9. Describe one advantage and one disadvantage of N-nucleotide addition during the rearrangement of immunoglobulin heavy-chain gene segments.

10. X-ray crystallographic analyses of many antibody molecules bound to their respective antigens have revealed that the CDR3 of both the heavy and light chains make contact with the epitope. Moreover, sequence analyses reveal that the variability of CDR3 is greater than that of either CDR1 or CDR2. What mechanisms account for the greater diversity in CDR3?

11. How many chances does a developing B cell have to generate a functional immunoglobulin light-chain gene?

12. Match the terms below (a–h) to the description(s) that follow (1–11). Each description may be used once, more than once, or not at all; more than one description may apply to some terms.

Terms

a. _____ RAG-1 and RAG-2
b. _____ Double-strand break repair (DSBR) enzymes
c. _____ Coding joints
d. _____ RSSs
e. _____ P-nucleotides
f. _____ N-nucleotides
g. _____ Promoters
h. _____ Enhancers

Descriptions

(1) Junctions between immunoglobulin gene segments formed during rearrangement
(2) Source of diversity in antibody heavy chains
(3) DNA regulatory sequences
(4) Conserved DNA sequences, located adjacent to V, D, and J segments, that help direct gene rearrangement
(5) Enzymes expressed in developing B cells
(6) Enzymes expressed in mature B cells
(7) Nucleotide sequences located close to each leader segment in immunoglobulin genes to which RNA polymerase binds
(8) Product of endonuclease cleavage of hairpin intermediates in Ig-gene rearrangement
(9) Enzymes that are defective in SCID mice
(10) Nucleotide sequences that greatly increase the rate of transcription of rearranged immunoglobulin genes compared with germ-line DNA
(11) Nucleotides added by TdT enzyme

13. Many B-cell lymphomas express surface immunoglobulin on their plasma membranes. It is possible to isolate this lymphoma antibody and make a high affinity, highly specific mouse monoclonal anti-idiotype antibody against it. What steps should be taken to make this mouse monoclonal antibody most suitable for use in the patient. Is it highly likely that, once made, such an engineered antibody will be generally useful for lymphoma patients?